INDIAN STATISTICAL INSTITUTE

Mid-semester Examination: 2014-2015 (First Semester)

M. Tech (QR & OR) II Year

Applied Stochastic Processes

Date: September 15, 2014 Full Marks: 60 Duration: $2\frac{1}{2}$ 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. (a) Define a compound Poisson process.
   (b) Let $\{M(t), t \geq 0\}$ be a compound Poisson process. Find the moment generating function of $M(t)$.
   (c) Hence find $E[M(t)]$.

[2 + 5 + 3 = 10]

3. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ and that U be a random variable having a uniform distribution on the interval [0, 2]. Calculate the mean of $N(t + U)$.

[5]

4. A person enlist subscriptions to a magazine, the number enlisted being a Poisson process with mean rate 9 per day. Subscribers may subscribe for 1 or 2 years independently of one another with respective probabilities $\frac{2}{3}$ and $\frac{1}{3}$. Suppose the commission earned by the person is a_1 for a 1-year subscription and a_2 for a 2-year subscription. Let $X(t)$ be the total commission earned by the person in $[0, t]$. Find $E[X(t)]$ and $\text{Var}[X(t)]$.

[5+2= 7]

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 distribution of T_1, T_2, ..., T_n.
   (b) Suppose $\lambda(t) = \lambda$ (a constant). Derive the maximum likelihood estimate of λ. Find a 95% confidence interval for λ.

[8 + (4 + 3) = 15]

6. Let $\{N(t), t \geq 0\}$ be a nonhomogeneous Poisson process with intensity function $\lambda(t)$. Find the pdf of T_1, the time at which the first event occur, given that $N(t) = 1$.

[5]

7. Consider a moving average process of order 2. Check whether the process is stationary. Derive the autocorrelation function of the process.

[6 + 2 = 8]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2014-15 (First Semester)

Course Name: M-TECH (QROR) II Year.

Subject Name: **Software Engineering**

Date: 15-09-2014 Maximum Marks: 100 Duration: 3 hours

(Answer all the questions)

1. State whether the following statements are true or false, explain your answer in brief. **[2 x 10 = 20]**
   a. In a software (industry level) development project coding takes maximum time.
   b. The software (industry level) product is completed before planned scheduled then its quality is good.
   c. Usability of software is a function requirement.
   d. Coding is the first step of in the software (industry level) development.
   e. Correctness is the fundamental characteristic of software.
   f. Modularity is a means of problem partitioning in software design.
   g. The goal of the design is to produce a structure where the modules have low level of cohesion and high coupling.
   h. There are three main views in the software architecture – module, component and connector.
   i. The fundamental approach of SE is to separate the development process from the products.
   j. Architecture of a software system impacts some of the key nonfunctional quality attributes like modifiability, performance, reliability, portability, etc.

2. Why software engineering is necessary? Write the names of any three process models and explain one of them in detail with its advantages and disadvantages. What are the characteristics of a good software product? **[3+(2+4+2+2)+5 = 18]**

3. What are the different views in the software architecture? Describe each of them in detail. What is the role of software architecture in a software system? How it is different from design? **[3+6+3+3 = 15]**

4. Is it reasonable to assume that if software is easy to test, it will be easy to maintain? Suppose that by putting extra effort in design and coding you increase the cost of these phases by 10%, but you reduce the cost of maintenance by 5%. Will you decide to put the extra effort, and why? **[4+6 = 10]**

5. Suppose a program for solving a problem costs X, and industrial level software for solving that problem costs 10X. Where do you think this extra 9X cost is spent? Suggest a possible breakdown of this extra cost with your reasons and justifications. **[10]**

6. Write a SRS (Software Requirement Specification) for any one the following: (a) Student registration system, (b) Library automation system. **[12]**

7. Find the Cost, Effort, Project duration, and People required using the intermediate semi-detached type COCOMO for the problem given below.
   CAD Co., Inc. wants to produce a system that will perform computer aided design for a home construction industry. The system will have seven modules and their size in lines of code is given in the brackets: (1) UICF (3KLOC), (2) 2DDGA (5.5KLOC), (3) 3DGA (6.5KLOC), (4) DBM (3.5KLOC), (5) CGDF (6KLOC), (6) PCF (3KLOC) and (7) DAM (7.5KLOC). The required model parameters are $a_i = 3.0$, $b_i = 1.12$, $c_i = 2.5$ and $d_i = 0.35$. The estimated effort multipliers for different cost drivers are as follows: data base size is high (1.08), product complexity is very high (1.30), main storage is very high (1.21), execution time constraints is high (1.11), programmer capability is very high (0.70), programming language experience is very low (1.07), modern programming practices is high (0.91), Use of software tools low (1.10), and all other related characteristics are normal. A review of the historical data indicates that the organizational average productivity for system of this type is 620 LOC/pm. Assume that the average salary of software engineers to be Rs. 60,000/- per month. **[3+6+3+3 = 15]**

------xxx------

Note: Books, note-books, computers, mobiles etc. are not allowed during examination. Only calculators can be used.

INDIAN STATISTICAL INSTITUTE

Mid-semester Examination on Reliability II

Maximum Marks: 50 **Time: 2 Hours** **Date: 16.09.2014**

1. List down the commonalities and differences between software and hardware reliability. [10]

2. If $\Pi(h,t) = (\pi_0(h,t), \pi_1(h,t),....)$ is the prior distribution of probabilities of number of unidentified bugs in a software with 'h' as history of outcomes and 't' indicating the number of trials conducted so far, find out the expression for one-step posterior probabilities and prove them. [10]

3. Develop Moranda's model on software reliability stating clearly the assumptions required for the model under consideration and give expressions for the parameters involved. If the data observed are time points when bugs are found out, do we get independent expressions for the parameters of the model? If no, what shall we do to solve for the parameters. [20]

4. Define likelihood ratio ordering of two probability distributions. [2]

5. Let us consider two priors $\pi^l(h_1, t_1)$ and $\pi(h_2, t_2)$ such that $\pi^l(h_1, t_1) \overset{LR}{>} \pi(h_2, t_2)$, where the notation indicates likelihood ratio ordering. Show that $\pi^l(h_1 s^r f^{K-r}, t_1+K) \overset{LR}{>} (h_2 s^r f^{K-r}, t_2+K)$ for any K. [10]

6. Consider $\Pi(h,t) = (0,1,0,....)$, then find out the likelihood ratio ordering of $\pi(hs,t+1)$ and $\pi(hf,t+1)$. [3]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination:2014

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

Subject Name: Advanced Statistical Methods

Date of Examination: 17.09.2014 Maximum Marks: 70 Duration: 2.5 hours

Note:

1. This paper carries 80 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 examples. [4]

2. Let y be a p-variate random variable. Assume $E(y) = \mu$ and $Cov(y) = \Sigma$.

   Show that. $E\left[(y-\mu)'\Sigma^{-1}(y-\mu)\right] = p$. Normality is not required. [4]

3. Suppose y is $N3(\mu, \Sigma)$, where $\mu = \begin{pmatrix} 3 \\ 2 \\ 5 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 6 & 1 & +3 \\ 1 & 12 & -4 \\ +3 & -4 & 3 \end{pmatrix}$

   a) Find the distribution of $z = 2y_1 - y_2 + 3y_3$.
   b) Find the joint distribution of $z_1 = y_1 + y_2 + y_3$ and $z_2 = y_1 - y_2 + 2y_3$.
   c) Find the distribution of y_2.
   d) Find the joint distribution of y_1 and y_3.
   e) Find the joint distribution of y_1, y_3, and $\frac{1}{2}(y_1 + y_2)$.
   f) Find a vector **z** such that $\mathbf{z} = (\Sigma^{1/2})^{-1}(\mathbf{y} - \mu)$ is $N_3(\mathbf{0}, \mathbf{I})$

   State the appropriate property in each case [1½ + 3 + ½ + 1 + 3 + 3 = 12]

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

$$\mu = \begin{bmatrix} 4 \\ 6 \\ 5 \\ 3 \end{bmatrix} \quad \text{and} \quad \Sigma = \left[\begin{array}{cc|cc} 3 & 0 & 0 & 4 \\ 0 & 7 & -3 & 0 \\ \hline 0 & -3 & 2 & 0 \\ 4 & 0 & 0 & 9 \end{array}\right]$$

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

5. Two samples are drawn from $N_2(\mu, \sum)$.

$$\text{Sample - 1}: \begin{bmatrix} 3 & 17 \\ 7 & 12 \\ 8 & 15 \\ 6 & 24 \\ 10 & 18 \end{bmatrix} \quad \text{and} \quad \text{Sample - 2}: \begin{bmatrix} 12 & 19 \\ 18 & 12 \\ 17 & 14 \\ 10 & 18 \end{bmatrix}$$

(a) Test the hypothesis H_0: $\mu_1 = \mu_2$. State the necessary assumptions

(b) If the null hypothesis is rejected test for each variable separately [8+4 = 12]

6. In a study involving 2 variables four groups with 50 cases in each group were studied.

The sample means for the four groups were $\bar{x}_1 = \begin{pmatrix} 10 \\ 20 \end{pmatrix}$, $\bar{x}_2 = \begin{pmatrix} 15 \\ 20 \end{pmatrix}$, $\bar{x}_3 = \begin{pmatrix} 20 \\ 25 \end{pmatrix}$, $\bar{x}_4 = \begin{pmatrix} 25 \\ 40 \end{pmatrix}$

The sample covariance matrices for each group were,

$$S_1 = \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix}, \quad S_2 = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}, \quad S_3 = \begin{bmatrix} 6 & 3 \\ 3 & 9 \end{bmatrix}, \quad S_4 = \begin{bmatrix} 12 & 2 \\ 2 & 9 \end{bmatrix}$$

The experimenter wished to test the hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ about the means of these variables:

Answer the following questions:

(a) Write down the underlying model and assumptions

(b)

1) What is the value of the Wilk's lambda?
2) What is the critical value of the significance test at the α level of .05?
3) What is the decision regarding the null hypothesis at the α level of .05?
4) Test for individual variables if required.
5) What are your conclusions?

(c)

1) What is the observed value of Roy's largest root test ?
2) What is the value of Hotelling - Lawley trace?
3) What is the observed value of the Pillai' test?

[3 +14 + 3 = 20]

7. 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

a) Write down the underlying model and the associated assumptions.
b) Write down the hypotheses that are to be tested.
c) Test the hypotheses.
d) Carry out univariate analyses.
e) 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 + 6 + 6 + 3 = 22]

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination: (2014 - 2015)

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

Subject Name: Database Management Systems

Date: September 18, 2014 **Maximum Marks: 50** **Duration: 2 hrs**

Answer all the questions.

1. Design an Entity-Relationship diagram for the following database management system. The database maintains information about staffs of a hospital, including doctors and nurses, and patients at the hospital. The information we need includes:

   - Staffs, including their names, addresses and permanent account numbers (PANs).
   - Patients, including their names, addresses, and the name of their insurance company.
   - Patients are each assigned to a ward (room).
   - Those staff who are nurses are assigned to zero or more wards. Each ward has at least one nurse assigned.
   - The staffs who are doctors are assigned to zero or more patients. Patients may or may not have a doctor assigned, and they may have more than one doctor. Patients in the same ward may have different doctors but will always have the same nurse(s).

   You may consider other attributes in the diagram. Mention clearly the assumptions you have made. [25]

2. Describe in details the architecture of a database management system. [25]

INDIAN STATISTICAL INSTITUTE
M. Tech. (QR & OR) 2nd YEAR
Year: 2014
MID SEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 19.09.2014 **Max. Marks: 100** **Time: 3 hours**

Answer any four from 1. to 6.

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

[2+3+5+5=15]

2. State the condition under which an LP solution will be an integer solution. State dual simplex algorithm to solve an LP and its usefulness in the context of finding integer solution. Define earliest start time, latest finish time and floats for critical path method.

[4+5+6=15]

3. Describe branch and bound algorithm to solve an integer programming problem. State the method of constructing Gomory's constraint to obtain integer solution. Formulate a k out n system as an optimization problem.

[6+5+4=15]

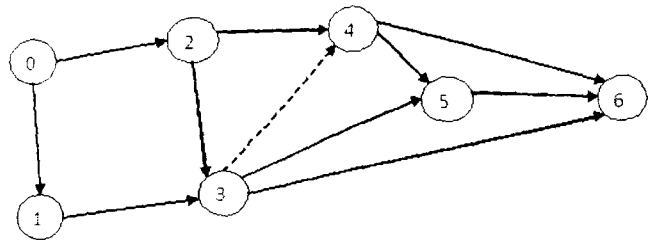
4. State Johnson's rule for n jobs, 2 machines and n jobs, 3 machines sequencing problem. Under what condition n jobs, 3 machines problem is solvable? Solve the following 2 jobs, 4 machines problem graphically.

	Machines			
Jobs	A	B	C	D
1	4	4	5	3
2	3	5	3	6

Technological ordering of Job-1 is A-B-C-D and of Job-2 is D-B-A-C.

[7+2+6=15]

5. Consider the network shown below:



The following table gives the estimated times for the different activities:

Activity (i, j)	Estimated Times (t_o, t_m, t_p)
(0,1)	(1,3,2)
(0,2)	(2,8,2)
(1,3)	(1,3,2)
(2,3)	(1,11,1.5)
(2,4)	(0.5,7.5,1)
(3,5)	(1,7,2.5)
(3,6)	(1,3,2)
(4,5)	(6,8,7)
(4,6)	(3,11,4)
(5,6)	(4,8,6)

Find the value of critical path. What is the variance of critical path? Find the probability that the project can be completed within 20 days (show all calculations and state your assumptions). The symbols t_o, t_m and t_p are optimistic, most likely and pessimistic time respectively.

[6+6+3=15]

6. State Bellman's principle of optimality for solving dynamic programming problem. Propose an iterative algorithm of multiplicative type to solve dynamic programming. Compare simplex algorithm and dual simplex algorithm.

[3+5+7=15]

7. Assignment

[40]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2014-15

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

Subject Name : Industrial Experimentation

Date: 22.09.2014 Maximum Marks: 100 Duration: 1 hours 40 minutes

NOTE: (i) This paper carries 107 marks. Question number 6 is compulsory. Answer as much as you can from the remaining 57 marks but the maximum you can score is 50. 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 any three of the following with suitable examples where ever feasible:

a) Industrial Experiment
b) Factor and Levels
c) Interaction between two factors
d) Blocking
e) Components of variance model

$(5 \times 3) = [15]$

2. Why is it said that "recognize the difference between practical and statistical significance"?

[5]

3. Consider a completely randomized design involving 'a' treatments. Under a fixed effects model, show that the $E(MS_{Treatments}) = \sigma^2 + n\frac{\sum_{i=1}^{a}\tau_i^2}{(a-1)}$.

[8]

4. A chemist wishes to test the effect of five chemical agents on strength of a particular type of cloth. Because there might be variability from one bolt to another, the chemist decides to use bolts of cloth as blocks in the experiment. He selects all the five available bolts and applies all five chemicals in random order to each bolt. The resulting tensile strengths are measured.

a) Write down the linear statistical model of the experiment, with necessary assumptions, if any.
b) Obtain the estimators of the model parameters using the method of least squares.

$(4 + 9) = [13]$

5. To improve the yield of a chemical process four different catalysts (A, B, C and D) are being studied. Each batch of material is only large enough to permit four runs to be made. Furthermore, each run requires approximately 1¾ hours, so only four runs can be made in one day. The experimenter decides to run the experiment as a Latin square, so that day and batch effects may be systematically controlled. The experiment was replicated. Analyse the following coded data (a higher value indicates a better yield compared to a lower value) from the experiment and draw appropriate conclusions.

	Replication							
	1				2			
	Day				Day			
Batch	1	2	3	4	5	6	7	8
1	$A = 8$	$B = 7$	$C = 10$	$D = 3$	$A = 9$	$B = 8$	$C = 10$	$D = 7$
2	$B = 4$	$A = 9$	$D = 1$	$C = 7$	$B = 5$	$A = 8$	$D = 3$	$C = 6$
3	$C = 11$	$D = 2$	$A = 7$	$B = 6$	$C = 10$	$D = 3$	$A = 8$	$B = 7$
4	$D = 6$	$C = 8$	$B = 3$	$A = 8$	$D = 4$	$C = 7$	$B = 3$	$A = 9$

[16]

6. Class and Home Assignments. [50]

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
Semester Examination: 2014-2015 (First Semester)

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

Applied Stochastic Processes

Date: November 24, 2014 Maximum Marks: 100 Duration: $3\frac{1}{2}$ hours.

Notes: (1) *This paper carries 110 marks. Answer as many questions as you can. The maximum you can score is 100.* (2) *Mention the basic results wherever necessary. Proofs are not required.*

1. Suppose that families migrate to an area in accordance with a Poisson process with rate $\lambda = 2$ per week. Let X_i be the number of members in the ith family. Suppose X_i's are independent and identically distributed random variables and takes on the values 1, 2, 3 with respective probabilities $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$. Also, X_i's are independent of the arrival process of the families. Find the mean and variance of the number of people migrating to this area during a fixed six-week period.

[10]

2. 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]

3. (a) Write down the postulates of a non-homogeneous Poisson process (NHPP).

   (b) Consider an NHPP $\{N(t), t \geq 0\}$ with intensity function $\lambda(t) = \alpha\lambda(\lambda t)^{\alpha-1}$. Find $E[\frac{N(t)}{t}]$.

[3+4=7]

4. (a) Define a covariance-stationary process.

   (b) Let W_0, W_1, W_2, ..., be uncorrelated with $E[W_n] = \mu$ and $\text{Var}(W_n) = \sigma^2$, $n \geq 0$. For some positive integer k, consider the process $\{X_n, n \geq k\}$ defined by

$$X_n = \frac{W_n + W_{n-1} + \cdots + W_{n-k}}{k+1}, \quad n \geq k.$$

   Is the process stationary? Justify your answer.

[2+7 =9]

5. Consider a Poisson process with rate λ. Show that given $N(t) = n$, the n arrival times $S_1 < S_2 < \ldots < S_n$ have the same distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $[0, t]$. [10]

6. A discrete time branching process starts with only one individual ($X_0 = 1$). The probability of producing k offspring by a single individual is p_k. For $0 < p_0 < 1$, prove that the eventual extinction probability (π) of this process is the smallest positive root of the equation $\pi = \phi(\pi)$, where $\phi(s) = \Sigma_{k=0}^{\infty} s^k p_k$.

[9]

P.T.O

7. (a) Let $X_1, X_2, \ldots$ denote the interarrival times of a renewal process with common expectation EX_1, such that, $0 < E(X_1) \leq \infty$. Let $M(t)$ be the number of renewals by time t. Prove that

$$\lim_{t\to\infty} \frac{M(t)}{t} = \frac{1}{EX_1} \text{ almost surely.}$$

(b) Suppose that passengers arrive at a train depot in accordance with a renewal process having a mean inter-arrival time μ. Whenever there are N passengers waiting in the depot, a train leaves. If the depot incurs a cost at the rate of nc rupees per unit time whenever there are n passengers waiting, what is the average cost incurred by the depot?

[5+4=9]

8. Consider a renewal process $\{X_i, i \geq 1\}$ with c.d.f. F. Let $\{U(t), t \geq 0\}$ denote the age process, i.e., $U(t) = t - Z_{M(t)}$ and $E(X_1) < \infty$.

(a) Show using the renewal reward theorem that, $\forall u \geq 0$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t P(U(\tau) \leq u) d\tau = \frac{1}{E(X_1)} \int_0^u (1 - F(x)) dx.$$

(b) Show by formulating a renewal equation for $U(t)$ that

$$\lim_{t\to\infty} P(U(t) > u) = \frac{1}{E(X_1)} \int_u^\infty (1 - F(x)) dx.$$

[4+8=12]

9. (a) Construct a birth and death process that is positive recurrent while its embedded Markov chain is null recurrent.

(b) Consider a birth-and-death process with $\lambda_{i,i+1} = \lambda$ for $0 \leq i \leq N-1$, $\mu_{j,j-1} = \mu$ for $1 \leq j \leq N$, and $\lambda_{ij} = \mu_{ij} = 0$ otherwise. $(0 < \lambda < \infty, 0 < \mu < \infty)$.

i. Show that this continuous time Markov chain is positive recurrent and find the invariant probability vector.

ii. Let n be a fixed number such that $0 < n < N$ and $M(t)$ be the number of $n \to n+1$ transitions in $[0, t]$. Obtain an expression for $\lim_{t\to\infty} \frac{1}{t} M(t)$.

[5+(3+8)=16]

10. (a) Consider a continuous time Markov chain with transition rate matrix Q and transition probability matrix $P(t)$. Prove that

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

(b) Calculate the transition probability matrix $P(t)$ for a Poisson process with rate λ.

(c) A single repair-person looks after both machines 1 and 2. Each time it is repaired, machine i stays up for an exponential time with rate $\lambda_i, i = 1, 2$. The repair time for machine i follows exponential distribution with rate μ_i, $i = 1, 2$. The repair-person will always service machine 1 when it is down. For instance, if machine 1 fails while 2 is being repaired, then the repair-person will immediately stop work on machine 2 and start on 1. Calculate the proportion of time machine 2 is down at steady-state.

[5+4+7=16]

INDIAN STATISTICAL INSTITUTE

First-Semester Examination: 2014-15

Course Name: M-TECH (QROR) II Year.

Subject Name: Software Engineering

Date: 24-11-2014 ***Maximum Marks: 100*** ***Duration: 3 hours***

Answer any five questions.

1. Explain the importance of Unified Modeling Language in detail. Why do we need to build model before building the actual system? Explain with example any three from the following in the context of Object oriented analysis and design:
   a) An Aggregation
   b) Inheritance and generalization
   c) Use case diagram
   d) State diagram

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

2. Define coupling and cohesion in object oriented (OO) paradigm. Explain in details the three types of Coupling and Cohesion, for each, in the OO paradigm. Explain any two relevant metrics for OO development paradigm.

   [4+ (2x5) + (2x3) = 20]

3. What is defect removal efficiency and how it is used in the software industry. Acceptance testing, System testing, Integration testing and Unit testing are different levels of testing, explain them in sufficient details.

   [4+4+4+4+4 = 20]

4. Explain in details the Halstead Complexity measures and the Cyclomatic Complexity measure with an example for each of them.

   [10 + 10 = 20]

P.T.O.

5. Explain the Alpha testing and Beta testing for software in details. Write down at least four relevant differences between them. Describe the advantages and disadvantages of Alpha and Beta testing.

[6+8+6 = 20]

6. Explain software quality? Define at least eight software quality attributes briefly.

[4+ (8 x 2) =20]

7. Explain software reliability. How is it different from Hardware reliability? Describe four metrics that help in measuring the software reliability.

[4+4+ (4x3) =15]

8. Write notes on the following (any four):
   a. Component based Software engineering (CBSE)
   b. Cleanroom Software Engineering
   c. Capability Maturity Model (CMM and CMMI)
   d. ISO/IEC for software
   e. Web Engineering
   f. Reengineering and Reverse-engineering

[4x5=20]

---XXX---

INDIAN STATISTICAL INSTITUTE
First Semester Examination: 2014-15

Course Name : M.Tech (QR & OR) II Year
Subject Name : Operations Research-II
Date: 26-11-2014 Maximum Marks: 100 Duration: 3 hours

Answer any five from (1) to (6).

1. (a) 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

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

(b) Define quasiconvex and pseudoconvex functions.

[10+6=16]

2. (a) Define co-positive and positive semidefinite matrices with distinct examples.

(b) Solve the following linear complementarity problem, LCP (q, M) by using complementary pivoting algorithm:

$$M = \begin{bmatrix} 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & -2 \\ 1 & -1 & 2 & -2 \\ 1 & 2 & -2 & 4 \end{bmatrix} \qquad q = \begin{bmatrix} 2 \\ 2 \\ -2 \\ -6 \end{bmatrix}.$$

[6+10=16]

3. (a) State the primal feasibility, dual feasibility and complementary slackness condition of a nonlinear programming problem.

(b) 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 $\lambda \in (0, \delta)$. Suppose that $f: R^n \to R$ is differentiable at $\bar{x}$. Prove that $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive semi-definite if $\bar{x}$ is a local minimum.

[9 + 7 = 16]

4. (a) Characterize the stationary point of a nonlinear programming problem in connection with the cone of feasible direction and the cone of descent direction.

   (b) State the KKT sufficient conditions of optimality for a nonlinear programming problem.

[8+8=16]

5. a) State the Fritz John necessary condition of optimality.

   b) Consider the problem:

$$\begin{aligned} \text{Minimize} \quad & (x_1 - 5)^2 + (x_2 - 3)^2 \\ \text{Subject to} \quad & 2x_1^2 + x_2^2 \le 7 \\ & x_1 + 3x_2 \le 4 \\ & x_1, x_2 \ge 0. \end{aligned}$$

   Write the KKT condition at (4,3).

[8+8=16]

6. (a) Formulate a separable programming problem as linear programming problem.

   (b) State a method to solve a linear fractional programming problem as linear programming problem.

[8+8=16]

7. Assignment

[20]

INDIAN STATISTICAL INSTITUTE

Semestral Examination : 2014-15

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

Subject : Reliability II

Date : 28.11.2014 Maximum Marks: 100 Duration : 3 hours

Notations used in this paper are usual notations used in the class. Answer all questions. This paper carries 100 marks. Marks for each question are given in [ ].

Q1.

a) Define the following classes of life distributions:
   i) DFR
   ii) DFRA
   iii) NWU
   iv) NWUE

b) Prove that DFR ⇒DFRA ⇒NWU ⇒ NWUE [2x4+12 = 20]

Q2.

Life of a certain group of items are assumed to follow exponential distribution with parameter λ. The following data are obtained:

i) The exact life of n items from the same group as $T_1, T_2, \ldots T_n$

ii) Data from a further m items whose life are known to be greater than M, but the exact lives are not known.

Calculate the maximum likelihood estimate of λ. [10]

Q3.

a) Write down clearly the steps involved in planning for an accelerated life test (ALT).

b) Classify the different models that are applicable for ALT.

c) A manufacturer of Bourdon tubes (used as a part of pressure sensors in avionics) wishes to determine its MTTE. The manufacturer defines the failure as a leak in the tube. The tubes are manufactured from 18 Ni (250) maraging steel and operate with dry 99.9% nitrogen or hydraulic fluid as the internal working agent. Tubes fail as a result of hydrogen embrittlement arising from the pitting corrosion attack. Because of the criticality of these tubes, the manufacturer decides to conduct ALT by subjecting

them to different levels of pressures and determining the time for a leak to occur. The units are continuously examined using an ultrasound method for detecting leaks, indicating failure of the tube. Units are subjected to three stress levels of gas pressures and the times for tubes to show leak are recorded. It is given that the time to leak in hours for each stress level follow Weibull distribution with the same shape parameter, but with different scale parameters.

The mean and 50th percentile at each stress level is given as follows:

Pressure (psi)	100	120	140
Mean	9276.1	4767.01	835.3
50th percentile	9050	4681	821

Determine the mean lives and the reliability functions for design pressures of 80 and 90 psi. [8+4+18 = 30]

Q4.

a) A system consists of four subsystems that must function if the system has to function properly. The system reliability goal is 0.950. All the four subsystems have identical reliability improvement effort functions. The estimated subsystem reliabilities at the present time are 0.75, 0.85, 0.90, 0.95. What reliability goal should be apportioned to the subsystems so as to minimize the total effort spent on the system improvement?

b) Using ARINC apportionment technique find out the allocation of reliability for each of the four components, assuming a mission time for the components and the system to be 20 hours

c) Give your comments, in case the allocated reliabilities are different in (a) and (b).

[10+8+2 = 20]

Q5.

a) Discuss the differences between software and hardware reliability. Describe at least two sources of uncertainty in software from the users' point of view.

b) State the assumptions of the Jelinski – Moranda model. Develop the model based on the assumptions and give expressions for the estimators of the unknown parameters when 'n' time between failures are noted as $t_1, t_2, \ldots, t_n$.

[(6+4)+(4+6) = 20]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2014-15

M. Tech. (QR & OR)-II

Industrial Experimentation

Date: 01.12.2014 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. Give answers to the following questions in single sentences not containing more than 20 words or by using appropriate phrases (one conjunction may be used per sentence):

a) What is a robust process?

b) Name the basic design principle used to improve the precision of comparisons among the factors of interest by reducing or eliminating the variability transmitted from known and controllable nuisance factors?

c) Is it appropriate to make conclusions about the levels tested for a design of experiments model where all factors are run at random (as opposed to fixed) levels?

d) What is the phrase used to call the process of examining the data to select comparisons of potential interest?

e) What is the name of a design of experiments in which all levels of a given factor are combined with all levels of every other factor in the experiment?

f) The difference in response between the levels of one factor is not the same at all levels of the other factors (e.g. the effect of factor A depends on the level of factor B). What is the name of this phenomenon?

g) What is the name of a two-factor design of experiments in which the levels of one factor (factor B) are similar but not identical for different levels of another factor (factor A)?

h) What is a 2^{k-p} design of experiments called where no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions?

i) How many factors and how many runs are there in a 2^{8-3} fractional factorial design?

j) What is the name given to the collection of mathematical and statistical techniques that are useful for the modelling and analysis of problems in which a response is influenced by several variables and the objective is to optimize this response?

k) In the analysis of a first-order response surface model, what is a path of steepest ascent?

$(2 \times 11) = [22]$

P.T.O.

2. Write short note on any three of the following:

a) Scheffé's method for comparing all contrasts,

b) Three basic principles of experimental design,

c) Minimum aberration design,

d) Addition of centre points to the 2^k design. $(6 \times 3) = [18]$

3. Show that $\hat{\tau}_i = kQ_i/(\lambda a)$ is the least squares estimators of the treatment effects in a BIBD having a treatments each replicated r times, arranged in b blocks of size k each where each pair of treatments appear in a block λ times. Q_i is the adjusted total for the ith treatment.

[12]

4. A 2^6 factorial experiment with factors A, B, C, D, E, F was performed in 8 blocks of 8 treatment combinations each. A particular block consisted of the treatment combinations

a, abcde, bd, abef, def, ce, acdf, bcf.

i) Obtain the principal block.

ii) What effects are confounded with blocks?

iii) If the experimenter can only use the contents of the principal block, is it possible to subdivide these into 4 smaller blocks of 2 units each so that no main effect is confounded with blocks? (*If your answer is 'no' then give justification, and if it is 'yes' then list these four smaller blocks. No credit will be given for answering just 'yes' or 'no'.*).

iv) Assume that the responses corresponding to the earlier given block (*a, abcde, ..., bcf*) are 71.9, 136.7, 94.1, 143.8, 87.3, 82.4, 73.4 and 91.3 respectively. If we have an independent estimate of mean square error as 87.31 based on 15 degrees of freedom then test for the significance of the possible effects by treating the given block as a 2^{6-3} fractional factorial design.

$(3 + 10 + 8 + 10) = [31]$

5. Discuss about the analysis of a second-order response surface, location of the stationary point and its characterisation as a minimum or a maximum or a saddle point based on canonical analysis of the fitted surface.

What is a central composite design (CCD)? What is rotatability of a CCD and how is it achieved?

$(12 + \overline{4 + 3}) = [19]$

6. a) Define Orthogonal Array. What is the role of Dr. G Taguchi's '*linear graph*' in designing fractional factorial experiments using such arrays?

b) At the manufacturing stage of carbon-resistance, carbon powder is coated on the surface of ceramic pipes. In order to find a better production condition an experiment is undertaken. In the study two different kinds of ceramic pipes (A) are considered for coating with two different kinds of carbon powder (B). Such coatings process requires heat and it is decided to examine four different carbon powder coating temperatures (C). After coating of the pipes with carbon powder, these are finished by providing a further coating of a non-conducting substance. Two finishing methods (D) – Enamel coating and Bakelite moulding, are selected for investigation.

For enamel coating, Kind of enamel (E) and Number of coatings (F) are suspected to influence the process. It is decided to study these two factors at <u>two levels each</u>. On the other hand for bakelite moulding <u>two</u> different moulding temperatures (G) are included for investigation.

Design the experiment in sixteen runs to study all the main effects and, two factor interactions $A \times D$, $B \times D$ and $C \times D$.

(2+3+13) = [18]

----- X -----

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

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

Col. No.	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

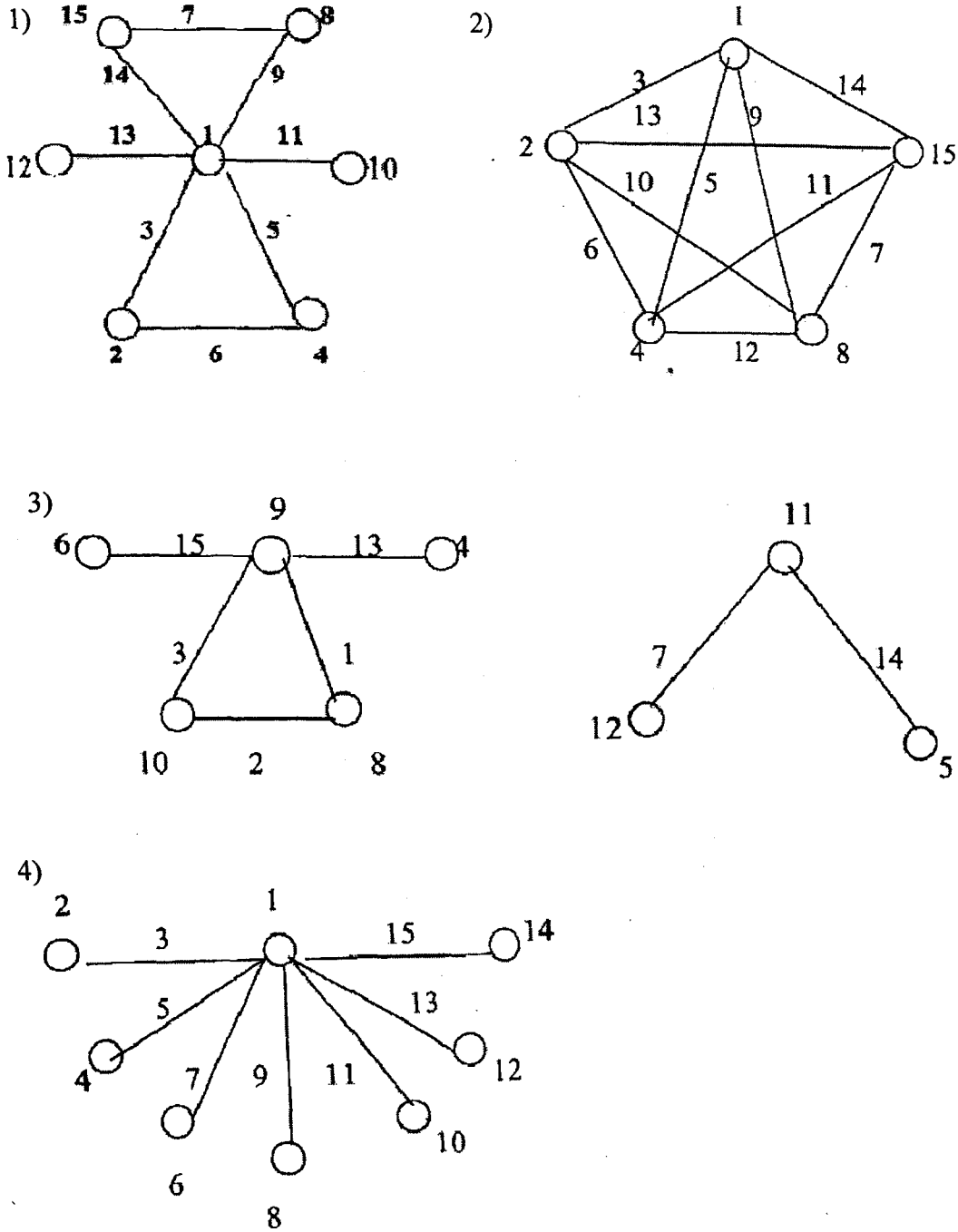


Fig. 1: Some Linear Graphs of OA(16,15,2,2)

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2014-15

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

Subject Name: Advanced Statistical Methods

Date of Examination: 05.12.2014 Maximum Marks: 100 Duration: 3½ hours

Note:
1. This paper carries 126 marks. Answer all questions but the maximum you can score is 100.
2. All notations have their usual meanings

---

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

2. Write True or False:
   i) Single linkage method and complete linkage method lead to the same cluster.
   ii) One may carry out several ANOVAs for each variable instead of a MANOVA. .
   iii) Principal components always lead to meaningful interpretation.
   iv) One can go back to the original variables in the case of principal component regression.
   v) A discriminant function always minimizes the distance between pairs of observations.
   vi) A model developed by multiple linear regression method never represents the underlying causal model.
   vii) A value of $R^2 = 0.8$ or above is enough to conclude that the underlying relationship is linear.
   viii) Factor analysis is an example of dependence technique.

   [8]

3. 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]

4. i) Identify the following models as a linear model, an intrinsically linear model or a non-linear model. Justify your answer. If a model is intrinsically linear, show how it can be linearised by a suitable transformation.

a) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

b) $y = \beta_0 x_1^{\beta_1} x_2^{\beta_2} x_3^{\beta_3} \varepsilon$

c) $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^{\delta} + \beta_3 x_3 + \varepsilon$

ii) Answer the following questions.

a) Write down the multiple linear regression model in matrix form

b) State the underlying assumptions

c) Write down the normal equations

d) Show that $\hat{\beta}$ (solution of normal equations) is unbiased for β

e) Derive the expression for V($\hat{\beta}$)

[4 + (2+3+2+4+3) = 18]

5. Consider a multiple linear regression model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots \beta_k x_k + \varepsilon$

i) Show that an equivalent way to perform the test for significance of overall regression in multiple linear regression is to base the test on R^2 as follows:

H_0: $\beta_1 = \beta_2 = \ldots\ldots = \beta_k = 0$ *versus* H_1: at least one $\beta_j \neq 0$, $j = 1,2 \ldots k$.

Calculate $F_0 = \dfrac{(n-k-1)R^2}{k(1-R^2)}$ and

Reject H_0 if the computed value of F_o exceeds $F_{\alpha, k, n-k-1}$.

ii) Suppose that a multiple linear regression model with k = 5 regressors have been fit with n = 40 observations and the R^2 was 0.85. Test the significance of the regression at $\alpha = 0.05$.

iii) What is the smallest value of R^2 that would have led to the conclusion of a significant regression?

[6 + 2 + 2 = 10]

6. 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

[2 + 5 + 5 + (3+3) = 18]

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

ii) Let X be a p x 1 random vector with dispersion matrix Σ and $Y_1, Y_2, \ldots Yp$ be the principal components, then show that

$$\sum_{i=1}^{p} V(Y_i) = \sum_{i=1}^{p} V(X_i)$$

iii) Let X be a random vector with dispersion matrix $\Sigma = \begin{bmatrix} 16 & 10 \\ 10 & 25 \end{bmatrix}$

Find the principal components. Calculate the proportion of total variability explained by the first principal component. Find the relative importance of X_1 and X_2 in determining the first principal component.

[3 + 6 + (10 + 3) = 22]

8. i) What are the similarities and dissimilarities between principal component analysis and factor analysis?

ii) Write down the orthogonal factor model and the associated assumptions. Explain the notations used.

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

iv) How does orthogonal transformation help in factor extraction?

v) Consider the three quality characteristics X_1 , X_2 *and* X_3. Based on a random sample of size n = 30, the following correlation matrix was obtained.

$$\boldsymbol{R} = \begin{bmatrix} 1.0 & 0.63 & 0.45 \\ 0.63 & 1.0 & 0.35 \\ 0.45 & 0.35 & 1.0 \end{bmatrix}.$$

Eigenvalues and corresponding eigenvectors of R are:

$\lambda_1 = 1.96,$ $e_1 =$ $(0.625,$ $0.593,$ $0.507)'$

$\lambda_2 = 0.68,$ $e_2 =$ $(-0.219,$ $-0.491,$ $0.843)'$

$\lambda_3 = 0.36,$ $e_3 =$ $(-0.749,$ $-0.638,$ $-0.177)'$

How many factors may be extracted? Calculate the Loading matrix.

[4 + 4 + 7+ 3 + 6 = 24]

9. i) What are the differences between hierarchical clustering and partitioning method of clustering?

ii) If we are to cluster a set of variables what is the suitable measure of dissimilarity?

iii) Suppose we measure two variables X_1 and X_2 for four items A, B, C and D. The data are as

Item	Observations	
	X_1	X_2
A	5	4
B	1	-2
C	-1	1
D	3	1

Use the K-means clustering technique to divide the items into K=2 clusters. Start with the initial groups (AB) and (CD).

[4 + 2 + 10 = 16]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2014 - 2015)

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

Subject Name: Database Management Systems

Date: December 03, 2014 **Maximum Marks: 100** **Duration: 3 hrs**

Answer as many questions as you wish. Maximum marks attainable is 100.

1. Consider the relation schema

   Inventory (*manufacturer*, *brandname*, *type*, *weight*, *store*).

   A relation under this schema stores the items that a grocery store stocks. Each tuple in the relation *Inventory* represents the fact that a store sells an item of a particular type and brand name manufactured by a particular company. The relation also stores the weight of the item. Two tuples that such a relation may contain are:

   (Kellogg's Company, Frosted Flakes, Cereal, 14oz., Hokies Holesome Foods) and
   (Kraft Foods, Philadelphia, Cream Cheese, 8oz., Healthy Hokies Store).

   Convert each of the two sentences (in i) and ii) below) in English about *Inventory* into functional dependencies. Consider each of these two sentences independently. In the following questions (Question 1.iii) - 1.ix) below), consider these two functional dependencies, if required.

   i) A manufacturer holds the trademark for a brand name of an item of a particular type, i.e., no two manufacturers can use the same brand name for items of the same type. (2)

   ii) For each type, each store sells only one brand name made by each manufacturer. (2)

   iii) What are the keys for *Inventory*? Justify your answer. (4)

   iv) What normal forms does *Inventory* satisfy? Justify. (5)

   v) Consider the decomposition of *Inventory* into

   Inventory1 (*manufacturer*, *brandname*, *type*, *store*) and
   Inventory2 (*manufacturer*, *brandname*, *type*, *weight*).

   Is it a lossless join decomposition? Justify. (2)

vi) If the decomposition in Question 1.v) is not lossless join, modify one of the attributes in either *Inventory1* or in *Inventory2* to obtain a lossless join decomposition. Justify your answer. (3)

vii) State with reasons all the normal forms that the decomposed relation schemas, obtained by you in your answer to Question 1.vi), satisfy. (7)

viii) Decompose *Inventory* into a set of relation schemas that are in BCNF. Is this decomposition dependency-preserving? (8)

ix) Obtain a lossless-join and dependency-preserving decomposition of *Inventory* into relation schemas that are in 3NF. Are all these relation schemas in BCNF? (7)

2. Consider the following relations schemas:
R_1 = (A, B, C, $\underline{D}$, E_1, D_1), foreign key E_1 references E of R_2, and foreign key D_1 references D of R_1;
R_2 = (G, $\underline{E}$, D_2), foreign key D_2 references D of R_1;
R_3 = (H, $\underline{J}$, E_1), foreign key E_1 references E of R_2;
R_4 = ($\underline{D_3}$, $\underline{J_3}$, K), foreign key D_3 references D of R_1, and foreign key J_3 references J of R_3.

i) Write an SQL query statement to make a list of all J-values for the tuples in R_3 that involve a tuple in R_1 with C = "c", associated with either a tuple of R_4 or a tuple of R_2. (10)

ii) Write the corresponding expressions in relational algebra, tuple relational calculus and domain relational calculus. (4 + 4 + 2 = 10)

iii) Obtain an optimized relational algebra expression, with respect to time and/or space, of the above query. (5)

3. Describe various mapping cardinalities, with appropriate real-life examples, of the Entity-Relationship model. (10)

4. Explain with appropriate examples various constraints that need to be incorporated while designing a database management system. (10)

5. Write short notes on any TWO of the following: (2 × 10 = 20)

i) Transaction management

ii) Storage & File structure, and Indexing & Hashing

iii) Query processing

iv) Hierarchical and Network models

v) Integrity and Security

vi) Object-oriented and Object-relational database models