INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2013-14 (First Semester)

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

Subject Name: Software Engineering

Date: 16-09-2013 *Maximum Marks: 60* *Duration: 2½ hours*

Note: The question paper is for 70 marks, answer as many as you can, you can get at most 60 marks.

1. State whether the following statements are true or false, if false correct the statement? [10 x 1= 10]

   (a) In a software development project coding takes maximum time.
   (b) If software is easy to test, it will be easy to maintain.
   (c) Usability of software is a function requirement.
   (d) Coding is the first step of the software development.
   (e) Correctness is the fundamental characteristic of software.
   (f) The structured Design Methodology (SDM) is primarily function-oriented design.
   (g) In design, Fan-in should be minimized and Fan-out should be maximized.
   (h) Maintainability and user-friendliness cannot be defined quantitatively.
   (i) The fundamental approach of SE to achieve the objectives is to separate the development process from the products.
   (j) Architecture of software system impacts some of the key nonfunctional quality attribute like modifiability, performance, reliability, portability, etc.

2. What are the different views of architecture? Describe each of them in detail. [10]

3. Define software engineering? What are the characteristics of good software? [10]

4. Explain the importance of project staffing and different staff structures along with their advantages. [10]

5. Use the structured design methodology to produce a design for a system that acts as a calculator with only basic arithmetic functions.

   Or

   Discuss some approaches on how you can use metrics to guide you in the design to produce a design that is easy to modify. [10]

6. Suppose that by putting extra effort in design and coding, you increase the cost of these phases by 14%, but you reduce the cost of maintenance by 8%. Will you put the extra effort in design and coding? Explain your reasons. [10]

7. Compute the Cost, Effort, Project duration, and People required using the intermediate semi-detached type COCOMO for the problem given below. [10]

   A Graphics Company wants to produce a system that will perform computer aided design for house construction industries. The system will have seven modules and their size in lines of code is given in the brackets: (1) Module-1 (2500), (2) Module-2 (5200), (3) Module-3 (6500), (4) Module-4 (3750), (5) Module-5 (3750), (6) Module-6 (2700) and (7) Module-7 (7700).
   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) and all other related characteristics are normal.

---XXX---

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

INDIAN STATISTICAL INSTITUTE
Mid-semester Examination: 2013-2014 (First Semester)

M. Tech (QR & OR) II Year

Applied Stochastic Processes

Date: September 16, 2013 Full Marks: 75 Duration: 3 hours.

Note: Answer all the questions

1. Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ.
   (a) Calculate E[N(t) N(t+s)].
   (b) Show that for $0 < s < t$,
   $$P[N(s) = k | N(t) = n] = \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}.$$
   (c) Let $S_1 < S_2 <, \ldots, < S_n$ be the occurrence time of n events of the process. Find the joint pdf of S_1, S_2, ..., S_n.
   (d) Derive the maximum likelihood estimate of λ based on occurrence time of n events. Find 95% confidence interval for λ.

[3+5+6+6=20]

2. Pulses arrive at a Geiger counter in accordance with a Poisson process at rate of 3 arrivals per minute. Each particle arriving at the counter has a probability 2/3 of being recorded. Let $X(t)$ denote the number of pulses recorded by time t minutes. Show that $X(t)$ is also a Poisson process. Find $E[X(t)]$.

[7+2=9]

3. Suppose $\{N(t), t \geq 0\}$ is a non-homogeneous Poisson process with intensity function $\lambda(t)$. Derive a differential equation for the probability generating function (pgf) of $N(t)$ and solve it. Hence find $P[N(t) = k]$.

[8+2 =10]

4. An insurance company pays out claims on its life insurance policies in accordance with a Poisson process having rate $\lambda = 2$ per week. If the amount of money paid on each policy is exponentially distributed with mean Rs.20000, find mean and variance of the amount of money paid by the insurance company in a four-week span? Assume that amount of claim and the claim process are independent.

[8]

P.T.O.

5. Consider a linear birth and death process $\{X(t), t \geq 0\}$ with equal birth and death rates [λ_n=μ_n, $n \geq 1$].

   (a) Find the pgf of $X(t)$ and hence find $E[X(t)]$.

   (b) Find the extinction probability.

[(8+3)+3=14]

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

   (a) Prove that

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

   (b) Find the transition probability matrix of the corresponding embedded Markov chain for

$$Q = \begin{bmatrix} -4 & 4 & 0 & 0 \\ 6 & -10 & 4 & 0 \\ 3 & 2 & -6 & 1 \\ 2 & 1 & 0 & -1 \end{bmatrix}.$$

   (c) Consider an $M/M/1/N$, where N is the system capacity (i.e. at most N customers can be accommodated by the system). Find the expected number of customers present in the system at steady state.

[6+2+6 =14]

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination: 2013-14

M. Tech (QR& OR) II year

Subject: Reliability II

Maximum Marks: 100 Duration: 3 hours Date:18.9.2013

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

1. (a) Define warranty.

(b) How it differs from guarantee?

(c) Classify different types of warranty?

(d) What is the difference between one dimensional and two dimensional non-renewing free replacement warranty policies?

(e) Explain the difference between basic rebate warranty policy and pro-rata rebate policy.

(f) Define with example (i) Two Dimensional Combination FRW/PRW
(ii) Cumulative Free Replacement Warranty
(iii) Reliability Improvement Warranty.

[2+1+3+4+5+(3+3+3) = 24]

2. (a) Explain Cox model.

(b) Why it is called semi-parametric proportional hazards model?

(c) If the covariates for the i^{th} life are $Z_i = (X_{i1}, X_{i2}, ..., X_{ip})$ and the vector of regression parameters is $\beta = (\beta_1, \beta_2, ..., \beta_p)$, give an expression for the constant ratio of the hazard of lives with covariate vectors Z_m and Z_n.

(d) Suppose there are patients some of whom are suffering from lung cancer. Following information is recorded for each patient.

$z_1 = 0$ for female
1 for male

$z_2 = 1$ for smoker
0 for non-smoker

$z_3 = 1$ for drinker
0 for non-drinker

$z_4 = 1$ for lung-cancer
0 for otherwise

Assume that, $\hat{\beta} = (0.031, 0.079, 0.033, 0.087)$.

Compare the force of mortality for a female non-smoker non-drinker non-lung-cancer patient with
(i) a male smoker, drinker, lung cancer patient.
(ii)a female smoker, non-drinker, lung cancer patient.
(iii)a male non-smoker, non-drinker, lung cancer patient.
(iv)a female non-smoker, drinker, lung cancer patient.

[2+3+2+10=17]

3. (a) Explain the expression for partial likelihood.

(b) Explain Breslow's approximation. When it is used?

(c) An investigation was carried out into the survival times (measured in months) of patients in hospital following kidney transplants. The covariates are
$z_{1i} = 0$ for Placebo
1 for treatment X
and z_{2i} =weight of patient (measured in kg).
The observed life times with weights in brackets are as follows:-

Placebo	Treatment X
3(87)	6*(59)
7(67)	11(72)
16(74)	17(66)
22(81)	24(80)
24(69)	24*(47)

Observations with * represents censored.
Using Breslow's approximation, find out what contribution to the partial likelihood is made by the deaths at time 24.

[2+3+4=9]

4. Consider the case of discrete software testing framework with $\pi(h,t)$ as the prior distribution of the number of undetected bugs and $g_k(\pi(h,t))$ denotes the net expected cumulative gain after testing 'k' test cases with $g_0(\pi(h,t)) = 0$ and 'h' is the history consisting of the outcomes of testing and 't' denotes the number of test cases tested. Assume that testing one test case will either result in a '*success*' when a bug is detected or a '*failure*' when no bugs are detected. A success gives 1 unit of reward whereas cost for testing a test case is 'c' where $c < 1$.
Under the above setup:

(a) Derive one step posterior distribution of the number of undetected bugs in the software.

(b) Show that if $\pi(h,t)$ has a finite mean so does the one step posterior distribution of the number of undetected bugs.

(c) Show that $g_k(\pi(h,t))$ converges to a limit and assume that the limit is $g(\pi(h,t))$.

(d) Define a reasonable decision rule for stopping time of the software testing using $g(\pi(h,t))$.

[4+7+7+2= 20]

5. Assume that $\pi(hfs, t+2) = \pi(hsf, t+2)$.
Show that the above condition leads to $q_n = q^n$ $(q = q_1)$ where q_n = Conditional probability of failure when there are n bugs in the software.

[4]

6. (a) Define likelihood ratio ordering of two probability distributions under the setup of Question 4.

(b) Prove that if $\pi'(h_1, t_1)$ is preferred in likelihood ratio ordering sense to $\pi(h_2, t_2)$, then the tail sum of the probabilities in the first distribution is greater than or equal to the tail sum of the probabilities in the second distribution.

[2+8=10]

7. Prove the following:

(a) For any distribution $\pi(h,t)$, $\pi(h,t) \overset{LR}{>} \pi(hf, t+1)$

(b)If $\pi'(h_1,t_1) \overset{LR}{>} \pi(h_2,t_2)$ then $\pi'(h_1 s, t_1+1) \overset{LR}{>} \pi(h_2 s, t_2+1))$ and $\pi'(h_1 f, t_1+1) \overset{LR}{>} \pi(h_2 f, t_2+1))$

[2+4=6]

8. (a) Describe clearly the assumptions of the Moranda's Geometric De-eutrophication model.

(b) Develop the model based on the above assumptions and find out expressions for the unknown parameters.

(c) Give your comments about how the solution can be implemented with actual data.

[3+6+1=10]

INDIAN STATISTICAL INSTITUTE

Mid Semestral Examination: (2013 - 2014)

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

Subject Name: Database Management Systems

Date: September 20 , 2013 **Maximum Marks: 50** **Duration: 2 hrs**

Answer all the questions.

1. Consider an insurance company database. This database could be modeled as the entity sets *EMPLOYEES*, with attributes *E#, name, and salary; SALESMEN*, with attributes *E#, name, and salary; MANAGERS*, with attributes *E#, name, and salary; and POLICIES*, with attributes *P#, policy-name, beneficiary* and *amount*. *SALESMEN* and *MANAGERS* are subsets of *EMPLOYEES*. Design an E-R diagram illustrating the relationships among these entity sets. Show the keys and the mapping cardinalities properly along with your assumptions. [20]

2. Describe the advantages of a database management system over a file processing system. [20]

3. Explain the following terms with appropriate examples. [4+4+4 = 12]
   i. DDL;
   ii. DML;
   iii. Atomicity property.

INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2013-14

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

Subject Name : Industrial Experimentation

Date: 23.09.2013 Maximum Marks: 70 Duration: 2 hours 15 minutes

NOTE: (i) This paper carries 76 marks. Answer as much as you can but the maximum you can score is 70. The marks are indicated in [ ] on the right margin.

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

1. What is an experiment? Mention some of the different objectives for which an experiment is undertaken. Write in one sentence the meaning of "the strategy of experimentation". What is meant by an interaction between two factors? How do you differentiate between replication and repetition? List the guidelines for designing an experiment (give the outline only, and do not elaborate).

(2+4+1+3+2+4) = [16]

2. Write down the fixed effects model along with necessary assumptions, for a balance and completely randomized design. Derive the least squares normal equations. Obtain the least square estimators of all the model parameters. Use general regression significance test procedure to obtain the treatment/factor sum of squares. Obtain the expected value of the error mean squares under the model. State the special form of the Cochran's theorem that is useful in establishing the independence of sum of squares due to error and sum of squares due to treatments.

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

3. What is a contrast? Write the expression for the sum of square due to a contrast when sample sizes are unequal? What is its degree(s) of freedom? When are two contrasts said to be orthogonal? Describe the Duncan's multiple range test for comparing all pairs of means.

(2+2+1+2+5) = [12]

4. An observation in a Latin square is missing. How do you propose to estimate it? Derive the formula for estimating the missing value based on your proposal. What is a standard Latin square?

   In a $p \times p$ Latin square, p batches of raw materials form the row blocks and p operators form the column blocks. The effect of five different ingredients (A, B, C, D, E) on reaction time of a chemical process is being studied. The experiment is replicated n times with new batches of raw material but with the same operators in each replicate. Set up the analysis of variance table for this experiment. (No derivation is required.)

(1+4+1+6) = [12]

5. An engineer is studying the mileage performance characteristics of five types of gasoline additives. In the road test he wishes to use cars as blocks; because of time constraint, he must use an incomplete block design. He runs the balanced design with the five blocks, randomly selected, that follow. Analyze the data from this experiment (use $\alpha = 0.05$) and draw conclusions.

	Car				
Additive	1	2	3	4	5
1	-	17	14	13	12
2	14	14	-	13	10
3	12	-	13	12	9
4	13	11	11	12	-
5	11	12	10	-	8

Obtain the intrablock, interblock and combined estimates of the treatment effects.

(10 + 6) = [16]

INDIAN STATISTICAL INSTITUTE

Mid-Semester Examination: 2013-2014

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

Subject Name: Advanced Statistical Methods

Date of Examination: 25.09.13 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. Let $\mathbf{Y} \sim N_4(\mu, \Sigma)$, where $\mu = \begin{bmatrix} 4 \\ 6 \\ 9 \\ 3 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 3 & 0 & 0 & 4 \\ 0 & 7 & -3 & 0 \\ 0 & -3 & 1 & 0 \\ 4 & 0 & 0 & 9 \end{bmatrix}$

   a) Write down the distribution of $Y_1 + Y_2 - 3Y_3 + Y_4$ in matrix form.

   b) Suggest a random variable that has $\chi^2_{(4)}$ distribution.

   c) Write down the distribution of $(Y_1, Y_4)'$.

   d) Write down the distribution of $(Y_1, Y_2, Y_4)'$.

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

   i) Y_1 and Y_2. ii) $(Y_1, Y_2)'$ and Y_3. iii) Y_2, Y_3 and Y_4.

   iv) $(Y_2, Y_3)'$ and Y_4. v) $(Y_2, Y_3)'$ and $(Y_1, Y_4)'$. vi) $(Y_1, Y_3)'$ and $(Y_2, Y_4)'$.

   f) What is the distribution of $\begin{pmatrix} Y_2 \\ Y_3 \end{pmatrix} - \begin{pmatrix} Y_1 \\ Y_4 \end{pmatrix}$

   g) What is the distribution of $\begin{pmatrix} Y_1 \\ Y_3 \end{pmatrix} + \begin{pmatrix} Y_2 \\ Y_4 \end{pmatrix}$

   [2 + 2 + 1 + 1 + 6 + 3 + 5 = 20]

2. Suppose Y is $N_3(\mu, \Sigma)$, where $\Sigma = \begin{bmatrix} 1 & \rho & 0.5 \\ \rho & 1 & \rho \\ 0.5 & \rho & 1 \end{bmatrix}$

   Is there a value of ρ for which $Y_1 + Y_2 + Y_3$ and $Y_1 - Y_2 - Y_3$ are independent? [8]

3. Consider the following problems:

a) In a steel melting shop there are two furnaces. The source of raw material is same for both the furnaces. Each day a single heat is produced and from each hea steel a single sample is collected and chemical analysis is carried out. Percentages four elements are recorded. A team of engineers has hypothesized that two furna produce steel of different chemical compositions.

b) Two batches of steel components have arrived. Physical properties (three in numb are measured taking samples from each batch. The manager wants to know if physical properties are the same for both the batches

Answer the following questions in each of case a) and b) above

i) Suggest the data collection plan that would facilitate testing the stated hypothesis
ii) Write down the hypothesis in each case
iii) Write down the necessary assumptions
iv) Write down the test statistic you propose to use in each case.
v) What distributions do they follow?
vi) What should be the minimum sample size in each case?
vii) What are the test criteria?
viii) Would you prefer univariate tests instead? Justify

[3 + 2 + 3 + 2 + 2 + 3 + 2 + 3 =

4. In a study to assess the effects of Solder-Bath Temperature (SBT) and Wave Height (WH) generation of defects in PCBs a 3^2 design was run. Three PCBs were soldered at each fac combination. Two types of defect were observed. Some intermediate computations are furnish 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.

Error Sums of Squares and Cross Products

	Defect 1	Defect 2
Defect 1	525.3333	438.33
Defect 2	438.33	11710.67

Between Effects Sums of Squares and Cross Products Matrices

Variable		Defect 1	Defect 2
SBT	Defect 1	246.2963	900.19
	Defect 2	900.19	5222.74
WH	Defect 1	2.0741	26.96
	Defect 2	26.96	55330.07
SBT*WH	Defect 1	49.4815	-89.52
	Defect 2	-89.52	9385.93

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

5. A study on evolution of the vocabulary of students was conducted on students of classes VIII, IX, X and XI. Vocabulary scores of 64 students were obtained. A sample of the data is given below.

Classes

Srl No.	VIII	IX	X	XI
1	1.75	2.60	3.76	3.68
2	0.90	2.47	2.44	3.43
3	0.80	0.93	0.40	2.27
4	2.42	4.15	4.56	4.21
5	−1.31	−1.31	−0.66	−2.22
6	−1.56	1.67	0.18	2.33
...	...	...	...	...
64	−0.04	2.44	1.79	2.64

It was felt that the vocabulary score of class XI may be predicted from a linear function of the vocabulary scores of the students in classes VIII, IX and X. A Computer output is given in the next page.

Write down the estimated a linear regression model and test its significance.

Computer Output:

A N O V A

	SS	df	MSS	F -test	p-value
Regression	166.150	3	55.383	47.386	0.0000
Residuals	70.126	60	1.169		
Total variation	236.276	63	3.750		

Multiple R	=	0.83857
R^2	=	0.70320
Adjusted R^2	=	0.68836
Standard Error	=	1.08110

PARAMETERS	Beta	SE	StandB	t-test	p-value
b [0] =	1.4579	0.3014	0.0000	4.838	0.0000
b [1] =	0.1974	0.1595	0.1925	1.238	0.2206
b [2] =	0.2265	0.1161	0.2439	1.952	0.0557
b [3] =	0.4042	0.1313	0.4535	3.079	0.0031

[10]

INDIAN STATISTICAL INSTITUTE
M. Tech. (QR & OR) 2nd YEAR
Year: 2013 - 14
MID SEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 27.09.2013 **Max. Marks: 100** **Time: 3 hours**

Answer any five.

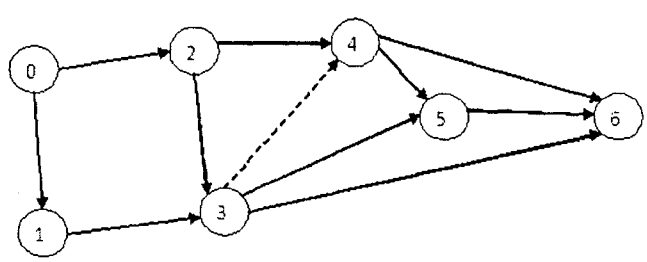
1. Describe Johnson's rule for an n jobs, 2 machines sequencing problem. Under what condition an n jobs, 3 machines problem is solvable? Solve the following 2 jobs, 4 machines problem graphically.

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

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

[7+5+8=20]

2. Consider the network shown below:



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

Activity (i, j)	Estimated Times (a,b,m)
(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 probability that the project can be completed within 20 days (show all calculations and state your assumptions). The symbols a, b and m have usual meaning.

[20]

3. Describe the cutting plane algorithm to solve an integer programming problem. State the method of constructing Gomory's constraint to obtain integer solution. How do you formulate a k out n system as an optimization problem?

[8+7+5=20]

4. A company is planning its advertising strategy for its three major products. The marketing department wants to establish how much to spend on each product in order to maximize the total sales. The following table gives the estimated increase in sales (in appropriate units) for each of the three different products for different advertising expenditures:

Advertising Expenditure	Product		
	1	2	3
1	5	6	4
2	7	8	9
3	12	10	11
4	16	12	15

Use dynamic programming to solve this problem.

[20]

5. 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.

[5+9+6=20]

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

[4+4+6+6=20]

INDIAN STATISTICAL INSTITUTE

First-Semester Examination: 2013-14

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

Subject Name: Software Engineering

Date: 2 - 12 - 2013 *Maximum Marks: 60* *Duration: 3 hours*

Answer any five questions

1. Explain software reliability. How is it different from hardware reliability? Describe any four metrics that help in measuring the software reliability.

[3 + 3 + 6 = 12]

2. What is UML? Explain the significance of UML. Why do we need to build model before building the actual system? Explain any two relevant metrics for Object Oriented (OO) development paradigm.

[1 + 3 + 4 + 4 = 12]

3. What is Cyclomatic complexity metric? Compute the Cyclomatric complexity of the following program.

```
0.  {
1.   i=1;
2.   while (i<=n) {
3.   j=1;
4.        while(j <= i) {
5.             If (A[i]<A[j])
6.             Swap(A[i], A[j]);        /* Swap(a,b) is a function for swapping values of a and b) */
7.              j=j+1;}
8.   i = i+1;}
9.  }
```

[4 + 8 = 12]

4. Write notes on the following (any three):
   a. ISO in Software Engineering.
   b. Web Engineering.
   c. Wideband Delphi estimation technique.
   d. Re-engineering and reverse engineering in Software Engineering.
   e. Refactoring in Software Engineering.

[4 x 3 = 12]

5. What is structured programming and how does it help improve code quality? What are the major concepts that help to make a program more readable?

[6 + 6 = 12]

6. Explain software quality? Define at least six software quality attributes in detail.

[3 + 9 = 12]

7. Explain black-box testing, glass-box testing, unit testing and system testing.

[3 + 3 + 3 + 3 = 12]

8. What is a Petri Net model in software engineering? Consider the traffic signal system on roads consisting of red, green and yellow lights with its usual meaning. Create a Petri Net model of traffic lights where the yellow light is on when lights are switching from red to green and back from green to red.

[4 + 8 = 12]

---XXX---

Note: Books, note-books, computers, mobiles etc. are not allowed during examination.

4. (a) Define current lifetime or spent lifetime ($Z(t)$) at time t. Prove that if $\mu < \infty$, then with probability 1,

$$\frac{Z(t)}{t} \to 0, \quad \text{as } t \to \infty.$$

(b) Describe age replacement and block replacement policies.

(c) A boy has an electronic gadget which works on a single battery. The boy does not keep any surplus batteries on hand. Each time a battery fails, the boy buys a new battery. The lifetime of a battery (in hours) is uniformly distributed over the interval (40, 60). The amount of time (in hours) it takes for him to get a new battery is uniformly distributed over (0. 1). Formulate this problem as a renewal process and find the expected renewal length. What is the average rate that the boy changes batteries?

[(2+5)+4+(5+2)=18]

5. 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. Let $M_D(t) = E[N_D(t)]$.

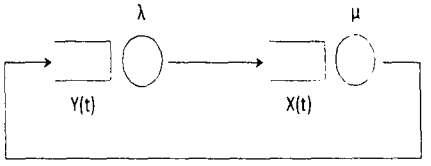
(a) Show that $M_D(t) = G(t) + \int_0^t M(t-x)dG(x)$, where $M(t)$ is the renewal function of ordinary renewal process with distribution F.

(b) Argue that the function $M_D(t)$ in (a) is the unique solution of the equation

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

[8+4=12]

6. (a) Two customers are trapped in a system comprising of two queues, as shown in the following figure. $\{X(t)\}$ and $\{Y(t)\}$ are the queue length processes as shown in the figure below. The service times in the two queues are exponentially distributed with parameters $\lambda > 0$ and $\mu > 0$. as shown. Argue that $\{X(t)\}$ and $\{Y(t)\}$ are CTMC and display their transition rate structures.



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

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

[4+6 =10]

7. 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 $\phi(s) = \sum_{k=0}^{\infty} s^k p_k$.

[10]

8. (a) Consider a moving average process of order 2. Calculate the autocorrelation function. Is the process stationary?- Justify your answer.

   (b) Consider a first-order autoregressive process:

   $$X_t = c + \phi X_{t-1} + Z_t, \quad \text{where } \{Z_t\} \sim WN(0, \sigma^2),\ c \text{ is constant and } |\phi| < 1$$

   Show that this process can be expressed as an MA(∞) process. Hence find the variance of X_t.

   (c) 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.
   Show that the process is not stationary. Show that the differenced process with appropriate order becomes an ARMA (p,q) process. Mention the values of p and q.

   [5+(3+2)+(2+5) = 17]

INDIAN STATISTICAL INSTITUTE

First Semester Examination: 2013-14

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

Subject Name : Operations Research-II

Date: 04-12-2013 Maximum Marks: 100 Duration: 3 hours

Answer any five from (1) to (7).

1. (a) State the geometrical aspects of finding the solution of LP and NLP problems.
   (b) Suppose A is an $m \times n$ matrix and c is an n dimensional vector. Then, show that exactly one of the following two systems has a solution:

   $System\ 1 \quad Ax < 0 \quad$ for some $x \in R^n$

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

   [7 + 9 = 16]

2. (a) Define convex function, pseudo-convex function and quasi-convex function.
   (b) Let $f(x_1, x_2) = 2x_1 + 2x_1^2 - 2x_1x_2 + x_2^2$. Find the Hessian matrix $H(x)$ and show that $H(x) \in$ Positive Definite.

   [9 + 7 = 16]

3. (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.

   [6 + 10 = 16]

4. (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]

5. 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]

6. (a) Define positive semidefinite matrix and copositive matrix.

Consider the matrix $A = \begin{bmatrix} 4 & 1 & 2 & 5 \\ 2 & 1 & 4 & 1 \\ 4 & 1 & 0 & -1 \\ 1 & 8 & 2 & 1 \end{bmatrix}$

Is the matrix A copositive-plus?

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

[6+4+6=16]

7. (a) Consider the following problem

Minimize $2x_1 - 6x_2 + x_1^2 + 2x_2^2$

Subject to $x_1 + 2x_2 \leq 12$

$x_1^2 + x_2^2 = 8$

$x_1 + x_2 \geq 0$

Make a suitable change of variables such that the problem becomes separable. Set up the initial simplex tableau to solve the approximated problem.

(b) Define sub-gradient of a function.

[12+4=16]

8. Assignment [20]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: (2013 - 2014)

Course Name: M. Tech. (QROR) **Year: 2nd year**

Subject Name: Database Management Systems

Date: December 06, 2013 **Maximum Marks: 100** **Duration: 3 hrs**

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

1. Draw an ER diagram that models the information in the following scenario: A manufacturing company has several assembly plants in different cities. Each plant produces one product which requires certain parts in its assembly. The parts are from appropriate suppliers, located in different cities. Certain parts may be used in more than one product. Identify appropriate attributes and show them in ER diagram. [20]

2. Consider a relational schema, R = {A, B, C, D, E}, with the set of functional dependencies, F = {B → CD, D → E, B → A, E → C, AD → B}. Answer with appropriate justifications. [2+2+2+2+2+2 = 12]

   i) Is B → E in F^+?

   ii) Is B a key for R?

   iii) Is D a key for R?

   iv) Is AD a key for R?

   v) Is AD a candidate key for R?

   vi) Is ADE a candidate key for R?

3. Write short notes on any FOUR of the following: [4×5 = 20]

   i) Transactions

   ii) Recovery System

   iii) Query Processing

   iv) Concurrency Control

   v) Indexing and Hashing

   vi) Storage and File Structure

   vii) Integrity and Security

4. Consider the following database of employees who work in specific departments. Each department has an inventory of items with specific quantity.

EMPLOYEE(PAN, first-name, last-name, address, date-joined, supervisor-PAN)
DEPARTMENT(dept-no, dept-name, manager-PAN)
WORKS-IN(employee-PAN, dept-no)
INVENTORY(dept-no, item-id, quantity)
ITEMS(item-id, item-name, type)

Foreign keys:
i) EMPLOYEE.supervisor-PAN, DEPARTMENT.manager-PAN and WORKS-IN.employee-PAN point to EMPLOYEE.PAN.
ii) WORKS-IN.dept-no and INVENTORY.dept-no point to DEPARTMENT.dept-no.
iii) INVENTORY.item-id points to ITEMS.item-id.

You are given below the relational algebra query.

$$R_1 = (\Pi_{\text{item-id}}\, \text{ITEMS}) - (\Pi_{\text{item-id}}\, \text{INVENTORY})$$

$$R_2 = \text{ITEMS} \bowtie_{\text{item-id=item-id}} R_1$$

$$R_3 = \sigma_{\text{type="CD"}}\, R_2$$

$$R_4 = \Pi_{\text{item-name}}\, R_3$$

i) Write an SQL query that returns the same answer as R_4 below. Explain your answer by writing the meaning of this query in English.
ii) Write the corresponding expression in Tuple Relational Calculus, the result of which is the same as that of R_4.
iii) Write the corresponding expression in Domain Relational Calculus, the result of which is the same as that of R_4. [12+10+8 = 30]

5. Consider the following relational schema of a patient database, Patient.

Patient (WardNo, WardName, Location, ChargeNurseName, ChargeNurseID, Telephone, PatientID, PatientName, DateonWaitingList, ExpectedStay, DateAdmitted, DateLeave, DateActualLeave, BedNumber)

The following functional dependencies need to hold on this relational schema:

PatientID DateAdmitted → DateonWaitingList ExpectedStay DateLeave DateActualLeave BedNumber},
PatientID → PatientName,
BedNumber → WardNo,
WardNo → WardName Location ChargeNurseID ChargeNurseName Telephone,
ChargeNurseID → ChargeNurseName,
ExpectedStay DateAdmitted → DateLeave.

a) Determine, with justifications, the normal formal which the above relational schema belongs to.
b) Decompose the above relational schema into the schemas that are in 3NF, if the original one is not in 3NF. Give reasons for each step of decomposition. State with reasons whether the functional dependencies remain preserved on this decomposition.
c) Decompose the given relational schema into the schemas that are in BCNF, if the given relational schema is not in BCNF. Give reasons for this decomposition. State with reasons whether this decomposition is dependency preserving.

[2+(10+2)+(8+2) = 24]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2013-14

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

Subject Name : Industrial Experimentation

Date: 09·12·13 Maximum Marks: 100 Duration 3 hours

NOTE: (i) This paper carries 118 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. Explain the terms/phrases:

main effect, interaction, blocking, random factor, minimum aberration design and crossed array design.

[2 + 2 + 3 + 2 + 4 + 4 = 17]

2. The response time in milliseconds was determined based on a completely randomized design, for three different types of circuits used in an automatic valve shutoff mechanism. The results are coded by subtracting a constant and are given below.

Circuit Type	Coded Response Time				
1	-4	-1	-3	-5	3
2	6	7	9	3	15
3	-5	-7	-3	5	-5

a) Write down the least square normal equations for this problem, and solve them for $\hat{\mu}$ and $\hat{\tau}_i$ considering the usual constraint ($\sum_{i=1}^{3} \hat{\tau}_i = 0$). Estimate $(\tau_1 - \tau_2)$ and its standard error given that the mean square error is 17.33.

b) Solve the equations in (a) using the constraint $\hat{\tau}_3 = 0$. Are the estimators $\hat{\tau}_i$ and $\hat{\mu}$ are the same as you found in (a)? Why? Now estimate $\tau_1 - \tau_2$ and compare your answer with (a). What statement can you make about estimating contrasts in τ_i's?

c) Estimate $\mu + \tau_1$, $2\tau_1 - \tau_2 - \tau_3$, $\mu + \tau_1 + \tau_2$, using the two solutions to the normal equations. Compare the results obtained in each case and comment.

[$\overline{3 + 1 + 2}$ + $\overline{3 + 1 + 2 + 2}$ + $\overline{3 + 3}$ = 20]

3. When are Latin square designs used to conduct experiments? What is a standard Latin square? When are two Latin squares orthogonal? How do you implement the principle of randomization in a Latin square design?

[2 + 1 + 3 + 4 = 10]

4. An industrial engineer is investigating the effect of three assembly methods (A, B, C) on the assembly time for a colour television component. Three operators are selected for the study. Furthermore, the engineer knows that helper assisting an operator also influences the assembly time. To account for this source of variability three helpers are selected for study. The engineer further suspects that the workplace use by the three operators may represent an additional source of variation. So as a fourth factor, he decides to include three different workplace layouts. Design the experiment. What is the type of design that you have obtained? What is the response being considered? Which one of the factors is treatment? Which is/are the factor(s) of interest and which is/are the factor(s) introduced to control the sources of extraneous variability?

[6 + 1 + 1 + 1 + 2 = 11]

5. To study the effect of heat treatment on the thickness of the carbonized layer of a gear, six two-level factors were investigated in a 2^{6-2} design: A = furnace temperature, B = cycle time, C = carbon concentration, D = duration of the carbonizing cycle, E = carbon concentration of the diffuse cycle, and F = duration of the diffuse cycle. The experiment is given below:

Standard Order	*A*	*B*	*C*	*D*	*E*	*F*	Standard Order	*A*	*B*	*C*	*D*	*E*	*F*
1	−	−	−	−	−	−	9	−	−	−	+	−	+
2	+	−	−	−	+	−	10	+	−	−	+	+	+
3	−	+	−	−	+	+	11	−	+	−	+	+	−
4	+	+	−	−	−	+	12	+	+	−	+	−	−
5	−	−	+	−	+	+	13	−	−	+	+	+	−
6	+	−	+	−	−	+	14	+	−	+	+	−	−
7	−	+	+	−	−	−	15	−	+	+	+	−	+
8	+	+	+	−	+	−	16	+	+	+	+	+	+

(a) Find the generating relations and the complete defining relation of the design. What is its resolution?

(b) Find the aliases of the effects A, AE and ABD.

(c) Is it possible to conduct the above experiment in two blocks of size 2^3 each, so that no main effect and no two factor interaction is confounded with the blocks? If your answer is yes then give the principal block and the associated defining contrast, and if it is no, then give justification.

$[\overline{8 + 2 + 1} + 3 + 6 = 20]$

6. (a) Explain the procedure followed to accommodate a two-level factor in a column of an appropriate three-level orthogonal array (OA).

(b) How do you account for the two degrees of freedom associated with a three-level column which contains a two-level factor?

(c) Consider the $L_9(3^4)$ array and give the decomposition of the eight degrees of freedom associated with the nine runs of the L_9 array containing only two two-level factors, A and B, in columns *1* and *2* respectively of the array.

(d) Explain how would you compute the $A \times B$ interaction sum of squares?

(e) Compute the error sum of square, using the following data: 1, 3, 6, 4, 8, 5, 2, 7, 0 for trial numbers 1, 2, .. , 9 respectively of the above OA for the allocation given in (c).

[3 + 3 + 4 + 4 + 6 = 20]

7. (a) When an arrangement of factors in a multifactor experiment is called a nested design?

(b) A manufacturing engineer is studying the dimensional variability of a particular component that is produced on three randomly selected machines. Each machine has two spindles, and four components are randomly selected from each spindle. The results of the experiment are given in the following table. Write the linear statistical model for this design with all the necessary assumptions. Analyse the data. Estimate the parameters of an appropriate model.

	Machine 1		Machine 2		Machine 3	
Spindle:	1	2	1	2	1	2
Data:	4	0	6	4	6	8
	1	1	7	2	2	7
	3	2	5	3	4	7
	4	0	6	5	3	6

$F_{0.05,2,2} = 19.00$, $F_{0.05,2,3} = 9.55$, $F_{0.05,2,18} = 3.55$, $F_{0.05,2,19} = 3.52$

$F_{0.05,3,18} = 3.16$, $F_{0.05,3,19} = 3.13$

$F_{0.01,2,2} = 99.00$, $F_{0.01,2,3} = 30.82$, $F_{0.01,2,18} = 6.01$, $F_{0.01,2,19} = 5.92$

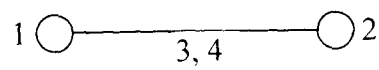
$F_{0.01,3,18} = 5.09$, $F_{0.01,3,19} = 5.01$

$[2 + \overline{3 + 9 + 6} = 20]$

Table: OA $L_9(3^4)$

No.	Col. 1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

Associated Linear Graph



INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2013-14 (First semester)

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

Subject Name: Advanced Statistical Methods

Date: 11 .12.2013 Maximum Marks: 100 Duration: 3½ hours

Note: This paper carries 113 marks. You can answer any part of any question, but maximum you can score is 100. All the notation have their usual meaning.

1) Write *Agree* or *Disagree* and briefly Justify

i) Single and complete linkage methods would always lead to the same cluster.
ii) Prior knowledge about the distribution of the variables is a prerequisite for clustering.
iii) Principal components do not 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 represents the underlying causal model.
vii) A multiple linear regression model cannot be used if it is hypothesised by the scientist that the response is related to the dependent variables as $y = \beta_0 x_1^{\beta_1} x_2^{\beta_2}$.
viii) One may carry out several ANOVAs for each variable instead of a MANOVA.

[2 x 8 =16]

2) (a) Show that the proportion of total population variance explained by i[th] principal component is

$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}$, $i = 1,...p$, where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ are the eigen values of dispersion matrix Σ of $X = (x_1\ x_2, \ldots x_p)'$.

(b) Consider the variance-covariance matrix Σ of $X = (x_1\ x_2, x_3)'$:

$$\Sigma = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

Find all the principal components. Is there any gain by extracting the principal components for this variance-covariance structure? Justify your answer.

(c) The weekly rates of return for five stocks (Allied Chemical, du-pont, Union Carbide, Exxon, and Texmaco) listed on a Stock Exchange were determined for a particular period. The data were collected for 100 successive weeks. The observations in successive weeks are assumed to be independently distributed, but the rates of return across stocks are correlated. The eigen values ($\lambda_1, \lambda_2, \ldots \lambda_5$) and corresponding normalised eigenvectors ($e_1, e_2, \ldots e_5$) of sample Correlation Matrix R were as follows.

$\hat{\lambda}_1 = 2.857, \quad \hat{e}_1 = (0.464, 0.457, 0.470, 0.421, 0.421)$

$\hat{\lambda}_2 = 0.809, \quad \hat{e}_2 = (0.240, 0.509, 0.260, -0.526, -0.582)$

$\hat{\lambda}_3 = 0.540, \quad \hat{e}_3 = (-0.612, 0.178, 0.335, 0.541, -0.435)$

$\hat{\lambda}_4 = 0.452, \quad \hat{e}_4 = (0.387, 0.206, -0.662, 0.472, -0.382)$

$\hat{\lambda}_5 = 0.343, \quad \hat{e}_5 = (-0.451, 0.676, -0.400, -0.176, 0.385)$

i) Describe two appropriate criteria for choosing the components.

ii) How many principal components should be retained in the present case?

[7 + (3+2) + (5+3) = 20]

3) a) Explain the difference between principal component analysis and factor analysis?

b) Obtain the principal component estimate of factor loadings. Show that the contribution due to the j^{th} factor to the total sample variance is λ_j, where $\lambda_1, \lambda_2, \ldots, \lambda p$ are the eigenvalues corresponding to S.

c) In a study to identify the factors, the following factor solution was obtained with respect to correlation matrix R:

Variables	Principal Component Loading		Varimax Rotated Loading	
	F_1	F_2	F_1^*	F_2^*
X_1	0.949	-0.295	0.884	0.455
X_2	0.974	-0.193	0.830	0.545
X_3	0.978	0.171	0.578	0.808
X_4	0.943	0.319	0.449	0.888

i) Calculate the communalities for each variable with respect to original loadings and rotated loadings.

ii) Are the communalities same before and after rotation? Why? Explain your answer.

iii) Calculate the contribution of each factor as proportion of total variance after rotation.

(4+10+(3+5+3)] = 25

4) (a) Consider two multivariate normal populations: $N_p(\mu_1, \Sigma_1)$ *and* $N_p(\mu_2, \Sigma_2)$. Derive the minimum Total Probability of Misclassification (TPM) rule to classify an observation x_0, assuming equal prior probabilities. Write down the expression of *Actual Error Rate* for the classification rule.

(b) Various aspects of economic cycles were measured for certain consumer goods and producer goods. The variables are x_1 = length of cycle, x_2 = percentage of rising prices, x_3 = cyclical amplitude and x_4 = rate of change. There were n_1 = 9 observations on consumers' goods and n_2 = 10 observations on producers' good. The following summary statistic were obtained:

$$\bar{x}_1 = \begin{pmatrix} 48.61 \\ 52.67 \\ 0.056 \\ 0.922 \end{pmatrix}, \quad \bar{x}_1 = \begin{pmatrix} 90.30 \\ 50.50 \\ 17.40 \\ 1.07 \end{pmatrix}, \quad S_{pooled}^{-1} = \begin{pmatrix} 0.00225 & 0.00227 & 0.00592 & 0.03793 \\ 0.00227 & 0.01723 & -0.01059 & 0.13159 \\ -0.00592 & -0.01059 & 0.08180 & -0.35849 \\ 0.03793 & 0.13159 & -0.35849 & 7.10212 \end{pmatrix}$$

$T^2(x_1, x_2, x_3, x_4) = 18.46248$, $T^2(x_2, x_3, x_4) = 11.64578$

$T^2(x_1, x_3, x_4) = 18.43656$ $T^2(x_1, x_2, x_4) = 15.06727$, and

$T^2(x_1, x_2, x_3) = 18.45909$.

i) Calculate the partial F-statistic corresponding to each variable, and find out the variables having significant contribution in group separation. Rank the variables in respect of relative contribution of group separation.

ii) Derive the linear discriminant function which maximally separates the groups. Which variable has highest contribution in group separation? Calculate the value of Mahalanobis distance.

iii) Consider an observation $x_0 = (66.5, 48, 15, 1.0)'$. Classify this observation as consumers' goods or producers' goods using the classification rule obtained in (a).

[12+ (6 + 6 + 6) = 30]

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

(b) What is the difference between hierarchical and partitioning method of clustering?,

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

$$\begin{bmatrix} 0 & & & & \\ 5 & 0 & & & \\ 6 & 8 & 0 & & \\ 10 & 9 & 5 & 0 & \\ 7 & 12 & 6 & 10 & 0 \end{bmatrix}$$

i) Cluster the five items using the complete-linkage method.

ii) Draw the dendogram and interpret.

[5 + 3 + (8 + 6) = 22]

INDIAN STATISTICAL INSTITUTE

Semestral Examination: 2013-14

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

Subject : Reliability II

Date: 13/12/2013. Maximum Marks: 100 Duration: 3 hours

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

1. (a) Consider that a device works only when its strength (X) is greater than its stress (Y). If reliability of the device is defined to be $P(X > Y)$, find out expressions for reliability, when the density function of X and Y are represented by $f_X(x)$ and $f_Y(y)$, respectively.

   (b) If both X and Y follow log-normal distribution with different parameters, find out the expression for reliability of the device.

   (c) The strength(X) and stress (Y) are log-normally distributed random variables with the following parameters:

$$\begin{aligned} E(X) &= 150000kP_a \\ E(Y) &= 100000kP_a \\ \sigma_Y &= 15000kP_a \end{aligned}$$

   Find out the minimum allowable standard deviation of the strength (X) so that the reliability does not fall below 0.990.

   (d) If both X and Y follow exponential distributions with λ_X and λ_Y as the respective parameters, show that the reliability of the device would be $\frac{\lambda_Y}{\lambda_X+\lambda_Y}$.

[4 + 12 + 10 + 4 = 30]

2. Define the following classes of life distributions:

(a) i) IFR and DFR,
ii) IFRA and DFRA,
iii) NBU and NWU,
iv) NBUE and NWUE.

(b) Show that the inequality in the definition of NBU becomes equal if and only if the life distribution is exponential.

(c) Show that for a life distribution F,

$$DFR \Rightarrow DFRA \Rightarrow NWU \Rightarrow NWUE$$

[8 + 6 + (3 + 6 + 3) = 26]

3. Show that if a device has n components and the components are independent with the lifetime distribution of each of the components follow IFRA, then the lifetime distribution of the device as a whole will have IFRA properties. [15]

4. Discuss the steps involved in planning for an accelerated life testing experiment. Also write the assumptions required to be made for using any accelerated life testing model. [5]

5. Suppose a system consists of four sub-systems that must function if the system has to function properly. The system reliability goal is 0.950.

(a) If we apply equal apportionment technique, what would be the reliability requirement of each subsystem in order to meet the system reliability goal?

(b) The estimated subsystem reliabilities presently are 0.75, 0.85, 0.90 and 0.95. If we apply ARINC apportionment technique, assuming that the mission time for the system as well as the subsystem is 20 hours, what will be the apportioned reliabilities of the subsystems in order to meet the given reliability goal of the system?

(c) Assume that all the four subsystems have identical reliability improvement effort functions. What reliability goal should be apportioned to the subsystem so as to minimize the total effort spent on the system improvement?

(d) Write down the conditions that the effort functions should satisfy in order to get an optimum solution to the problem?

[3 + 8 + 10 + 3 = 24]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : 2013-14

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

Subject Name : Industrial Experimentation

Date: 12/03/2014 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. Fill in the blanks with appropriate ***phrase*** or ***word***:

a) Three basic principles of experimental design are _____, _____, _____. A _____ is a set of relatively homogeneous experimental conditions. [1 · 4 4]

b) A resolution III fractional factorial design is said to be saturated if _____ [2]

c) The block containing treatment combination (I) is called the _____ block. [1]

d) If an interaction contains at least one random effect and one fixed effect then the entire interaction is considered as _____. [1]

e) The highest possible resolution for a 2^{15-11} design is _____. [1]

f) A 2^{7-4} fractional factorial design with two replications provides information about _____. [2]

g) In a multifactor experiment, if the levels of one factor are similar but not identical for different levels of another factor then such an arrangement is called a _____ design. [1]

h) A design that minimizes the number of words in the defining relation that are of minimum length, amongst the designs having same resolution, is called a _____ design. [2]

i) We usually take as the path of steepest ascent, the line through the _____ of the region of interest and _____ to the fitted surface. [1 · 2 2]

j) Generally a central composite design consists of a 2^k _____ (or _____ design of resolution ___), _____ _____ runs and a few _____ runs. [1 · 7 7]

k) A single experimental design containing both controllable and noise factors is usually called a _____ _____ design. [2]

2. What is a randomised complete block design? Write the appropriate statistical model for such a design. Derive the normal equations and solve them under the usual constraints ($\sum_{i=1}^{a} \hat{\tau}_i = 0$) and ($\sum_{j=1}^{b} \hat{\beta}_j = 0$). Under a fixed effects model, write down the expressions for expected mean squares for treatments and for blocks.

[3 + 3 + 3 + 3 + 2 = 14]

3. Consider the 2^6 full factorial design, confounded in eight blocks of eight runs each with *ABCD*, *ACE* and *ABEF* as the independent effects chosen to be confounded with blocks. Generate the principal block. Find other effects confounded with blocks.

[8 + 4 = 12]

4. Write short notes on ***any three*** of the following:
   a) Duncan's multiple range test.
   b) Graeco-Latin square design.
   c) Intra-block analysis of a BIB design.
   d) Split-plot design.
   e) Method of steepest ascent.

(7 × 3) = [21]

5. Discuss the response surface approach to a robust design problem using two controllable factors x_1 and x_2, a single noise factor z_1 and an appropriate *first-order* model.

[7]

6. An experiment was run using two two-level factors: arsenic flow rate (*A*) and deposition time (*B*). Two replicates were run and the epitaxial layer thickness (in some coded units) was measured. The factor levels and the coded data are shown below.

		Replicate			Factor Levels	
A	B	I	II		Low (−)	High (+)
−	−	-4.2	-4.8	A	55%	59%
+	−	-3.6	-4.8			
−	+	4.6	4.9	B	Short	Long
+	+	0.3	5.4		(10 min)	(15 min)

(a) Estimate the effects.

(b) Conduct an analysis of variance. Which effects are important?

(c) Write down a regression equation that could be used to predict epitaxial layer thickness over the region for factors studied in the experiment.

(d) Use the regression model in part (c) to find the setting of arsenic flow rate and deposition time required to obtain a layer thickness of 1.1 units.

(e) How would your analysis of variance table change if arsenic flow rate is more difficult to change from trial to trial and, no more than four runs can be conducted in a day?

[Use the *F* distribution table given overleaf.] (5 + 4 + 3 + 3 + 6) = [21]

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
Back Paper Examination: 2013-2014 (First Semester)

M. Tech (QR & OR) II Year

Applied Stochastic Processes

Date: 13 March, 2014 Full Marks: 100 Duration: 3 hours.

Notes: (1) Answer all questions. (2)Symbols have their usual meanings.

1. A radioactive source emits particle in accordance with a Poisson process with parameter λ. Each particle emitted has a probability p of being recorded and not recorded with probability $1-p$. Let $N_1(t)$ denote the number of particles recorded and $N_2(t)$ the number of particles not recorded by time t. Derive the distributions of $N_1(t)$ and $N_2(t)$. Find the expectations of $N_1(t)$ and $N_2(t)$.

[10+2=12]

2. Suppose $\{N(t), t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t)$.

   (a) Write down the postulates for the process.

   (b) Derive a differential equation for the pgf of $N(t)$ and solve it. Hence find $P[N(t) = k]$.

   (c) Suppose $\lambda(t) = 0.2t^{-1/2}$. Derive the pdf of T_1, the time at which the fist event occurs.

[2+(7+3)+4 =16]

3. Consider a repairable system, which is observed until the eleventh failure. The failure times in days are

$$9, 20, 65, 88, 104, 107, 138, 143, 149, 186, 208.$$

   Assume that the failure process can be modelled by a homogeneous Poisson process with parameter λ.

   (a) Derive the maximum likelihood estimate of λ.

   (b) Derive a 95% confidence interval for λ.

[5 + 7 = 12]

4. Consider a renewal process with interarrival time distribution F. If F is uniform [0, 1], prove that

$$M(t) = e^t - 1, \quad 0 \leq t \leq 1.$$

[10]

5. 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]

6. Let $\{N(t), t \geq 0\}$ be a renewal process having interarrival times X_n, $n \geq 1$ and suppose that each time a renewal occurs we receive a reward. Let R_n be the reward earned at the time of the nth renewal. Assume that the R_n, $n \geq 1$, are independent and identically distributed. Let $R(t)$ represents the total reward earned by time t. Then show that if $E[R_1] < \infty$ and $E[X_1] < \infty$,

$$\lim_{t \to \infty} \frac{R(t)}{t} = \frac{E[R_1]}{E[X_1]}, \quad \text{with probability 1,}$$

[10]

7. Consider a general birth and death process $\{X(t), t \geq 0\}$ with birth and death rates λ_n $(n \geq 0)$ and μ_n $(n \geq 1)$, respectively. Derive the expression of equilibrium distribution. Write down the condition for existence of the equilibrium distribution.

[8+2=10]

8. 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]

9. Let $Z_0, Z_1, Z_2, \ldots$ be uncorrelated random variables with $E[Z_n] = 0, n \geq 0$ and

$$\text{Var}(Z_n) = \begin{cases} \sigma^2/(1-\lambda^2) & \text{for } n = 0 \\ \sigma^2 & \text{for } n \geq 1, \end{cases}$$

where $\lambda^2 < 1$.
Define $X_0 = Z_0$ and $X_n = \lambda X_{n-1} + Z_n, \quad n \geq 1$.

Is the process $\{X_n, n \geq 0\}$ stationary? Justify your answer.

[8]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination: 2013-2014

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

Subject Name: Advanced Statistical Methods

Date of Examination: 14.03.2014 Maximum Marks: 100 Duration: 3 hours

Note:
1. Answer all questions
2. All notations have their usual meanings

---

1. Write True or False:

   a) Single linkage method & complete linkage method lead to the same cluster.
   b) Prior Knowledge of the distribution of the variables is a prerequisite for clustering.
   c) Principal components do not always lead to meaningful interpretation.
   d) One can go back to the original variable in the case of principal component regression.
   e) A discriminant function always minimizes the distance between pairs of observations.
   f) A model developed using multiple regression technique never represents the underlying causal model.

   [1 X 6 = 6]

2. An experiment was conducted to study the relationship of Yield (**y**) of a synthetic analogue to jojoba oil in terms of x_1: Reaction Temperature, x_2: Initial Amount of Catalyst and x_3: Pressure. Following table gives the results

Expt No.	x_1	x_2	x_3	y
1.	-1	-1	-1	17
2.	1	-1	-1	44
3.	-1	1	-1	19
4.	1	1	-1	46
5.	-1	-1	1	7
6.	1	-1	1	55
7.	-1	1	1	15
8.	1	1	1	41
9.	0	0	0	29
10.	0	0	0	28
11.	0	0	0	30
12.	0	0	0	27
13.	0	0	0	28

   a) Fit a multiple linear regression model relating **y** to x_1, x_2 and x_3, and test for lack of fit.
   b) Is it possible to get the contribution of each of the regressors unconditionally on the others? Explain your answers?

   [12 + 5 = 17]

3. A chemical engineer wants to compare two types of protective coatings (say A and B) for resistance to corrosion. 15 pipes each were coated using coating type A and B respectively. 15 pairs of pipes (one coated with type A and the other with type B) were formed and buried at 15 different locations and left for the same length of time.

Corrosion was measured as follows:

y_1: Maximum depth of pit (in thousandths of an inch) for type A

y_2: Number of pits for type A

x_1: Maximum depth of pit (in thousandths of an inch for type B

x_2: Number of pits for type B

a) State and test the appropriate hypothesis.

b) Carry out univariate tests.

c) What are your conclusions?

Maximum depth of Pits and Number of Pits of coated pipes

Location	Coating A		Coating B		Difference	
	Depth	Number	Depth	Number	Depth	Number
	y_1	y_2	x_1	x_2	$d_1 = y_1 - x_1$	$d_2 = y_2 - x_2$
1	73	31	51	35	22	-4
2	43	19	41	14	2	5
3	47	22	43	19	4	3
4	53	26	41	29	12	-3
5	58	36	47	34	11	2
6	47	30	32	26	15	4
7	52	29	24	19	28	10
8	38	36	43	37	-5	-1
9	61	34	53	24	8	10
10	56	33	52	27	4	6
11	56	19	57	14	-1	5
12	34	19	44	19	-10	0
13	55	26	57	30	-2	-4
14	65	15	40	7	25	8
15	75	18	68	13	7	5

[12 + 5 + 3 = 20]

4. In a study to assess the effects of Solder-Bath Temperature (SBT) and Wave Height (WH) on generation of defects in PCBs a 3^2 design was run. Three PCBs were soldered at each factor combination. Two types of defect 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.

Error Sums of Squares and Cross Products

Variable	Defect 1	Defect 2
Defect 1	525.3333	438.33
Defect 2	438.33	11710.67

Between Effects Sums of Squares and Cross Products Matrices

	Variable	Defect 1	Defect 2
SBT	Defect 1	246.2963	900.19
	Defect 2	900.19	5222.74
WH	Defect 1	2.0741	26.96
	Defect 2	26.96	55330.07
SBT*WH	Defect 1	49.4815	-89.52
	Defect 2	-89.52	9385.93

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

5. Obtain the Fisher's discriminant function for more than two populations. [7]

6. i) Discuss the differences between Hierarchical clustering and Partitional algorithms.
ii) What are Scree plots and Dendograms? [5 + 5 = 10]

3. A chemical engineer wants to compare two types of protective coatings (say A and B) for resistance to corrosion. 15 pipes each were coated using coating type A and B respectively. 15 pairs of pipes (one coated with type A and the other with type B) were formed and buried at 15 different locations and left for the same length of time.

Corrosion was measured as follows:

y_1: Maximum depth of pit (in thousandths of an inch) for type A

y_2: Number of pits for type A

x_1: Maximum depth of pit (in thousandths of an inch for type B

x_2: Number of pits for type B

a) State and test the appropriate hypothesis.

b) Carry out univariate tests.

c) What are your conclusions?

Maximum depth of Pits and Number of Pits of coated pipes

Location	Coating A		Coating B		Difference	
	Depth	Number	Depth	Number	Depth	Number
	y_1	y_2	x_1	x_2	$d_1 = y_1 - x_1$	$d_2 = y_2 - x_2$
1	73	31	51	35	22	-4
2	43	19	41	14	2	5
3	47	22	43	19	4	3
4	53	26	41	29	12	-3
5	58	36	47	34	11	2
6	47	30	32	26	15	4
7	52	29	24	19	28	10
8	38	36	43	37	-5	-1
9	61	34	53	24	8	10
10	56	33	52	27	4	6
11	56	19	57	14	-1	5
12	34	19	44	19	-10	0
13	55	26	57	30	-2	-4
14	65	15	40	7	25	8
15	75	18	68	13	7	5

[12 + 5 + 3 = 20]

4. In a study to assess the effects of Solder-Bath Temperature (SBT) and Wave Height (WH) on generation of defects in PCBs a 3^2 design was run. Three PCBs were soldered at each factor combination. Two types of defect 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.

Error Sums of Squares and Cross Products

Variable	Defect 1	Defect 2
Defect 1	525.3333	438.33
Defect 2	438.33	11710.67

Between Effects Sums of Squares and Cross Products Matrices

	Variable	Defect 1	Defect 2
SBT	Defect 1	246.2963	900.19
	Defect 2	900.19	5222.74
WH	Defect 1	2.0741	26.96
	Defect 2	26.96	55330.07
SBT*WH	Defect 1	49.4815	-89.52
	Defect 2	-89.52	9385.93

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

5. Obtain the Fisher's discriminant function for more than two populations. [7]

6. i) Discuss the differences between Hierarchical clustering and Partitional algorithms.
ii) What are Scree plots and Dendograms? [5 + 5 = 10]