INDIAN STATISTICAL INSTITUTE

Mid-Semestral Examination : 2016-17

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

Subject Name : Industrial Experimentation

Date: 19/09/2016 Maximum Marks: 65 Duration: 2 hours

NOTE: (i) This paper carries 75 marks. Answer as much as you can but the maximum you can score is 65. The marks are indicated in [ ] on the right margin.

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

1. a) Define an experiment.
   b) What is "strategy of experimentation"?
   c) When do you say two factors interact?

   [2+2+2 = 6]

2. a) **List** the seven guidelines for designing an experiment.
   b) Why is it said that "use your nonstatistical knowledge" for experimentation?

   [7+4 = 11]

3. Write the formula for computation of $SS_{Treatment}$ under a fixed effects model, for an unbalanced completely randomized design (CRD), describing the notations used and find the $E(MS_{Treatment})$.

   [1+5 = 6]

4. What is a contrast? What are orthogonal contrasts? Describe the Scheffe's method for comparing all contrasts.

   [2+2+5 = 9]

5. What is the design used to study a set of treatments after eliminating heterogeneity in one direction? Write down the model stating appropriate assumptions. Obtain the least squares estimators of your model parameters. How will you obtain the residuals (not the residual sum of squares) for your design?

   [1+3+5+2 = 11]

6. What is a Graeco-Latin square?
   Five treatments (A, B, C, D, E) are studied in the presence of three nuisance factors which are controlled during experimentation. The following data are obtained.

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

Analyse the data. Which treatment / treatments will you recommend if the response is a larger-the-better characteristic? [Give objective justification for your answer.]

[10+5 = 15]

7. What is a BIB(a, b, r, k, λ) design? Show that $\lambda(a-1) = r(k-1)$. Write down the model and obtain the normal equations. Solve these equations to obtain least squares estimators of the treatment effects. Show that these are unbiased estimators of treatment effects.

[2+2+4+4+5 = 17]

F distribution (5%) Table

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

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

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

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

PERCENTAGE POINTS OF THE T DISTRIBUTION

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

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

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

Maximum Marks = 100 Time = 3 hrs. Date : 20 September 2016.

1. Suppose the stress acting on a component is uniformly distributed over an interval $[S_{min}, S_{max}]$. The strength of the component has a two-parameter Weibull distribution with parameters S_0, and θ. Derive an expression for the reliability of the component. Let

$$S_{min}=10, \quad S_{max}=30$$
$$S_0=20, \quad \theta=30.$$

[20+5=25]

2. A system consists of six subsystems that must function if the system has to function properly. The system reliability goal is 0.990. All the six subsystems have identical reliability improvement effort functions. The estimated subsystem reliabilities at the present time are 0.70, 0.75, 0.80, 0.85, 0.90, 0.95. Determine what reliability goal should be apportioned to the subsystems so as to minimize the total effort spent on the system improvement?
If we apply the ARINC apportionment technique with the system mission time as 20 hours and the present reliability of the subsystems calculated for 50 hours, determine the apportioned reliabilities of the subsystems in order to achieve the system reliability goal of 0.990.

[15+15=30]

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

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

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

[5+2+8=15]

4. Let the prior distribution $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ of the undetected number of faults be given as {0.01, 0.02, 0.03, 0.04, 0.2, 0.3, 0.2, 0.1, 0.05,0.05, 0, 0, 0,

P.T.O

...}. Consider that the conditional probability of success given that there were i bugs in the software is denoted by $p_i=\frac{\pi i}{20}$. Find out the two-step-posterior probability π_2 (ss,2).

[10]

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

[20]

INDIAN STATISTICAL INSTITUTE

Mid – Semester Examination: 2016 - 17

Course Name: M Tech (QROR)

Subject Name: Business Analytics

Date: 21 September 2016 Maximum Marks: 100 Duration: 3 hours

Notes, if any: The paper has 5 questions. Answer any four.

1. Explain the following
   a. Supervised and unsupervised analytics. Give two examples of each. [4 + 4 = 8]
   b. What is the meaning of "over fitting"? What impact does an over fitted model has on the quality of prediction? [4 + 2 = 6]
   c. What are predictive and explanatory analytics? [3 + 3 = 6]
   d. What is meant by flexibility of a model? Do you think that a model is more likely to be over fitted as its flexibility increases? Why? [2 + 3 = 5]
2. Answer the following
   a. Name three conditions that make least squares a "good method". Explain how residual plots may be used to verify whether these conditions hold. [6 + 6 = 12]
   b. Describe briefly the meaning of leverage with an example. [5]
   c. Suppose you are trying to fit a multiple linear regression model where two variables are nominal with 3 and 4 levels each. How will you incorporate these variables in the model? Explain the concept briefly. [5]
   d. What is multi collinearity in multiple linear regression? What measures would you use to detect the presence of multi collinearity? [2 + 1 = 3]
3. Suppose you are using a decision tree for the purpose of estimating average sale in a retail store given certain conditions.
   a. Write down the model. [3]
   b. Briefly sketch the approach undertaken to fit the decision tree. [6]
   c. Suppose you have 20 variables used to fit the tree.
      i. How many terminal nodes can a fully grown tree have? [2]
      ii. If you fit a fully grown tree, do you envisage any difficulty? Explain briefly. How will you solve this problem? [3 + 5 = 8]
   d. Explain the concept of boosting a regression tree and sketch the algorithm. [6]
4. Suppose an email service provider wants to develop a spam filter. Suppose they have noted 10 different conditions [e.g. presence of exclamation mark (!), presence of certain specific words etc.] that are often observed in spam mails. You have been allocated the job and you have decided to use naïve Bayes' classifier.
   a. Explain the naïve Bayes' classifier in this context. How many parameters do you need to estimate? [10 + 2 = 12]
   b. Suppose you want to compare whether naïve Bayes' classifier would work better than random forest. How would you use the concept of leave one out cross validation (LOOCV) to take this decision? Explain in brief. [6]
   c. In this context explain the advantages and disadvantages of holdout samples vis-à-vis leave one out cross validation (LOOCV)? Explain briefly. [7]

5. Suppose you are fitting a logistic regression model to identify which variables contribute towards the occurrence of a particular disease. Note that the response is the status (0 or 1) of the person being studied, where 0 indicates the person would not have get the disease. Suppose further, you have 10 explanatory variables X_1, X_2, ...X_{10}. Explain the following in this context
    a. Write down the logistic regression model in terms of the response variable (say Y) and the explanatory variables. [4]
    b. Suppose the logistic regression is being used to assess the risk of a person getting the disease. Explain the concept of sigmoid curve and the role of logistic regression in this context. While explaining "risk" provide a definition of the same. [3 + 2 + 2 = 7]
    c. Suppose the estimated value of the coefficient of X_2 is 1.16 and the same is found to be significant (p value < 0.001). What is the interpretation? [3]
    d. Suppose you have fitted a logistic regression and it has been reported that the deviance divided by the degrees of freedom is 1.73. Do you think it indicates any problem? If yes, why? How will you address the same? [3 + 4 = 7]
    e. Can the logistic regression be always used to estimate probability of certain events? Explain. [4]

INDIAN STATISTICAL INSTITUTE
M. Tech. (QR & OR) 2nd YEAR
Year: 2016
MID SEMESTER EXAMINATION

Subject: Operations Research-II

Date of Exam: 22.09.2016 **Max. Marks: 100** **Time: 3 hours**

Answer any Five.

1. Define the following concepts of game theory:
   i. Pure and Mixed strategy.
   ii. Pay-off matrix.
   iii. Two person zero-sum game.
   iv. Equilibrium or Saddle point.
   v. Solve the game whose pay-off matrix is given below:

$$\begin{bmatrix} 30 & 40 & -80 \\ 0 & 15 & -20 \\ 90 & 20 & 50 \end{bmatrix}$$

[3+3+3+3+8=20]

2. Describe the branch and bound algorithm to solve an integer programming problem. State the method to formulate an integer programming problem into binary integer programming.

[15+5=20]

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

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

[20]

P.T.O

4. Formulate the following game problem as linear programming problem and solve using simplex algorithm:

$$\begin{bmatrix} 1 & -1 & 3 \\ 3 & 5 & -3 \\ 6 & 2 & -2 \end{bmatrix}$$

[20]

5. (a) Obtain the optimal strategies for both players and the value of the game graphically whose payoff matrix is given as below:

$$\begin{pmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 6 \\ 4 & 1 \\ 2 & 2 \\ -5 & 0 \end{pmatrix}$$

(b) Use matrix oddment method to solve the following game:

$$\begin{pmatrix} 4 & 7 \\ 6 & 5 \end{pmatrix}$$

[14+6=20]

6. Describe all integer 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?

[10+5+5=20]

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2016
Course Name: M. Tech. (QR & OR) 2nd YEAR
Subject Name: Advanced Multivariate Analysis

Date of Examination: 23.09.16 Maximum Marks: 70 Duration: 2 hours

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

---

1) Distinguish between *dependence* and *interdependence* techniques with at least two examples each [6]

2) Let $\mathbf{Y} \sim N_4(\mu, \Sigma)$, where

$$\mu = \begin{bmatrix} 4 \\ 6 \\ 9 \\ 3 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 3 & 5 & -8 & 4 \\ 5 & 7 & 0 & 0 \\ -8 & 0 & 1 & -2 \\ 4 & 0 & -2 & 9 \end{bmatrix}$$

Find the

(i) distribution of $Y_1 + Y_2 - 3Y_3 + Y_4$

(ii) joint distribution of $Y_1 - 2Y_4 + Y_3$ and $Y_1 - 3Y_2 + 2Y_3$

(iii) joint distribution of Y_1, Y_3 and Y_4

(iv) joint distribution of Y_3, Y_4 and $(Y_1 + Y_2 + Y_3)$

State the necessary result in each case.

$(2 + 4 + 2 + 4 = 12)$

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 thousands of an inch for type A

y_2: Number of pits for type A

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

x_2: Number of pits for type B

(i) State and test the appropriate hypothesis.

(ii) Carry out univariate tests

(iii) 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

[10 + 4 + 3 = 17]

4) (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 + (8 + 3) = 20]

5) 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.

(i) Write down the underlying model and the associated assumptions.

(ii) Write down the hypotheses that are to be tested.

(iii) Test the hypotheses.

(iv) Carry out univariate analyses if necessary.

(v) 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 + 5 + 5 + 3 = 20]

INDIAN STATISTICAL INSTITUTE
Mid-Semestral Examination : 2016-17
M. TECH. (QR & OR) II YEAR
Applied Stochastic Processes and Time Series Modelling

Date: 23 September 2016 Full Marks: 85 Duration: 3 hours

Answer all questions. Notation have usual meaning.

1 Suppose that exactly one event of a Poisson process has taken place by time t. Then, prove that the distribution of the time at which the event occurred is $U(0, t]$.

[6]

2 A shop is open from 10:00 to 16:00. Customer arrivals to the shop is governed by Poisson process with time dependent rate described in the following table:

Period	Rate/hour
10:00-12:00	6
12:00-14:00	15
14:00-16:00	linearly decreases from 15 to 10

(a) Write down the intensity function of the arrival process.

(b) Find the probability that no customer enter the shop between 13:00 and 15:00.

[4+6 = 10]

3 State and prove the Chapman-Kolmogorov equation for continuous-time Markov chain. **[10]**

4 Consider a birth-death process with state-space $S = \{0, 1, 2, 3, \}$, and the rates are $\lambda_n = (3 - n)^2$ and $\mu_n = n^2 + n$. Write down the generator of the process. Hence or otherwise give the Kolmogorov forward equations in matrix form.

[5+5 = 10]

5 Customers arrive at a store in groups consisting of 1 or 2 individuals with equal probability, and arrivals of groups are in accordance with Poisson process with rate λ per minute.

(a) Derive the expected number of customers arrived at the store in t minutes.

(b) If $\lambda = \frac{1}{2}$ per minute, what is the probability that total number of customers arrived in 4 minutes is exactly 4?

[6+8 = 14]

[P.T.O.]

6 We define

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

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

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

(b) Find $\mathrm{P}[N(2) - N(1) > 1 \mid N(1) = 1]$.

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

[5+4+6 = 15]

7 (a) Suppose $\{X_t\}$ is strictly stationary and $E(X_t^2) < \infty$ for all t. Then show that $\{X_t\}$ is weakly stationary.

(b) Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. Calculate $\mathrm{Corr}(N(t_1), N(t_2))$ for $t_1, t_2 \geq 0$.

(c) Consider a time series model

$$X_t = a_0 + a_1 t + Z_t$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and, a_0 and a_1 are constants. Show that $\{X_t - X_{t-1}\}$ is a stationary time series.

[8+5+7 = 20]

———*** xXx ***———

INDIAN STATISTICAL INSTITUTE
Mid-Semester Examination: 2016-17
Course Name: M.Tech (QR & OR), 2nd YEAR (E & S Streams)

Subject Name: Six Sigma

Date: 23/09/2016 **Maximum Marks: 100**

Duration: 2 hrs.

Answer All Questions

1. Quality is a state in which value entitlement is realized by customer and provider in every aspect of the business relationship. Here 'value' is a function of:

   a) loss incurred to the society after shipment of the product
   b) economic worth of the product
   c) economic worth, form, fit and function of the product
   d) economic worth, practical utility and availability of the product

   [2]

2. Reduction of variation around the target of a process leads to:

   a) reduction of defects
   b) reduction of societal loss after shipment of the product
   c) reduction of societal loss, defect and cycle time of a product
   d) reduction of societal loss and cycle time of a product

   [2]

3. Cost of poor quality becomes directly proportional to the profit margin of a company through:

   a) reduction of external failure costs only
   b) reduction of internal failure costs only
   c) reduction of appraisal and failure costs
   d) a substantial reduction of appraisal and failure costs by putting more emphasis on prevention costs

   [2]

4. An ideal Six Sigma project should have the desirable features as:

   a) critical to customer
   b) complex; requiring team approach
   c) 3-6 months duration
   d) The above three and a high bottom-line impact.

   [2]

5. In a Six Sigma project the important process variables are determined from:

   a) voice of the customer (VOC)
   b) critical to business requirements (CBR)

c) CTQs based on VOC/CBR
d) Customer survey

[2]

6. SIPOC provides a high-level view of a process that helps to:

a) apply in all kinds of work, whether repetitive in nature or one-of-a-kind
b) define project boundaries (starting and ending points) and describe where to collect data
c) none of the above
d) all of a and b

[2]

7. Identify the "odd-element-out" from the following elements of a team charter

a) project description and scope
b) expected business results
c) expected bottlenecks
d) team members

[2]

8. In a Pareto analysis if number of defects are translated into corresponding monetary loss

a) a large number of defects generally represent a great amount of money lost
b) a large number of defects may not represent a great amount of money lost
c) a small number of defects always represent a great deal of money lost
d) none of the above

[2]

9. Select the correct equation from the following:

a) $Z_{LT} = Z_{ST} + 1.5$
b) $Z_{LT} = Z_{ST} - 1.5$
c) $Z_{LT} = Z_{ST} \times 1.5$
d) $Z_{LT} = Z_{ST} \div 1.5$

[2]

10. The Headquarters of a bank assigns a Six Sigma team to investigate the deposit process and reduce the number of complaints at the branches. The bank provides the practitioners with 280 deposit slips (units). A collection of data after each step revealed the following:

- Customer gives deposit slip, cash and checks to bank teller (5 defects)
- Bank teller (associate) processes customer's deposits (6 defects)
- Bank Associate enters deposit in Bank's electronic system (3 defects)

- System generates deposit confirmation document (4 defects)
- Customer receives deposit confirmation document from Bank Associate (2 defects)

The overall sigma level of the deposit process is:

a) 1.47
b) 2.47
c) 2.97
d) 3.97

[5]

11. The Headquarters of a bank assigns a Six Sigma team to investigate the deposit process and reduce the number of complaints at the branches. The bank provides the practitioners with 280 deposit slips (units). A collection of data after each step revealed the following:

- Customer gives deposit slip, cash and checks to bank teller (5 defects)
- Bank teller (associate) processes customer's deposits (6 defects)
- Bank Associate enters deposit in Bank's electronic system (3 defects)
- System generates deposit confirmation document (4 defects)
- Customer receives deposit confirmation document from Bank Associate (2 defects)

The benchmark sigma level of the deposit process is:

a) 2.19
b) 3.69
c) 2.96
d) 3.96

[5]

12. $\%R\ \&\ R = \frac{R\ \&\ R}{TV} \times 100$. Find the correct relational equation from the following:

a) $R\ \&\ R = \sqrt{TV^2 + PV^2}$
b) $PV = \sqrt{TV^2 + R\ \&\ R^2}$
c) $R\ \&\ R = \sqrt{TV^2 - PV^2}$
d) $TV = \sqrt{R\ \&\ R^2 - PV^2}$

Note: R&R implies repeatability & reproducibility; PV stands for part variation and TV represents total study variation.

[2]

13. In any study concerning measurement system adequacy, if it is found that %R&R is less than 10% then the decision is

a) MS is acceptable
b) May be acceptable based on importance of application and cost of measurement
c) MS is unacceptable
d) None of the above

[2]

14. In a schematic Box Plot the outliners can be defined as:

a) points that lie beyond the upper and lower adjacent values
b) points that lie below the upper adjacent value
c) points that lie above the lower adjacent value
d) point that lie below the upper adjacent value but above the lower adjacent value

[2]

15. Box Plot is a five-number summary of a set of data. The five numbers are:

a) Lowest value, first quartile, mean, third quartile, highest value
b) Lowest value, first quartile, mode, third quartile, highest value
c) Lowest value, first quartile, median, third quartile, highest value
d) None of the above

[2]

16. Variation in a process exists due to:

a) Chance causes
b) Assignable causes
c) Chance and assignable causes
d) None of the above

[2]

17. Which of the following statements are TRUE about the six-sigma philosophy?

1. Sigma is a statistical term, which measures how far a process varies from perfection
2. Quantitatively, Six Sigma is a stretched goal of achieving 3.4 defects per million opportunities
3. Six Sigma is a term used to describe any cost saving achieved by the management
4. Culturally, Six Sigma is about learning to build processes that deliver defect free output continuously

a) Statement 1 only
b) Statements 2 and 3 only
c) Statements 1, 2 and 4
d) All of the above

[5]

18. For a customer-driven organization, the supplier-customer relationship can best be described using

a) IPO diagram
b) SIPOC diagram
c) Affinity diagram
d) COPIS diagram

[2]

19. For effective Six Sigma implementation, the number of employees trained should be:

a) 1 champion per business group, 1 MBB per 30 BBs, 1 BB per 10000 employees, 1 GB per 20 employees
b) 1 champion per business group, 1 MBB per 30 BBs, 1 BB per 1000 employees, 1 GB per 20 employees
c) 1 champion per business group or manufacturing site, 1 MBB per 30 BBs, 1 BB per 1000 employees, 1 GB per 100 employees
d) 1 champion per business group or manufacturing site, 1 MBB per 30 BBs, 1 BB per 100 employees, 1 GB per 20 employees

[2]

20. Which of the following is not a part of the project charter?

i. Problem statement
ii. Goal statement
iii. Proposed solutions
iv. Project CTQs
v. Constraints of project CTQs
vi. Project scope
vii. Project schedule
viii. Project Teams
ix. Control plans
x. Business case

a) i. & ii.
b) iv., v. & x.
c) vi., vii. & viiii.
d) iii. & ix.

[5]

21. A stable, normally distributed process with specification 3.50 ± 0.05 has average of 3.50 and standard deviation of 0.025. What percent of the production violates specification?

a) 4.55%
b) 5.55%
c) 5.46%
d) 5.00%

[5]

22. Which of the following is NOT an expected output of the "Define" phase?

1. Clear objectives for the project
2. Scope and deliverables for the project
3. CTQ-KPIV Correlation
4. Cause & Effect diagram
5. Roles of team members
6. Data Collection Plan

a) 1 , 2 and 3
b) 2, 4, and 6
c) 3 ,4 and 6
d) 5 and 6

[5]

23. At Six Sigma performance level, one expects:

a) Max 2.4 DPMO
b) Min 3.4 DPMO
c) Max 4.3 DPMO
d) Max 3.4 DPMO

[2]

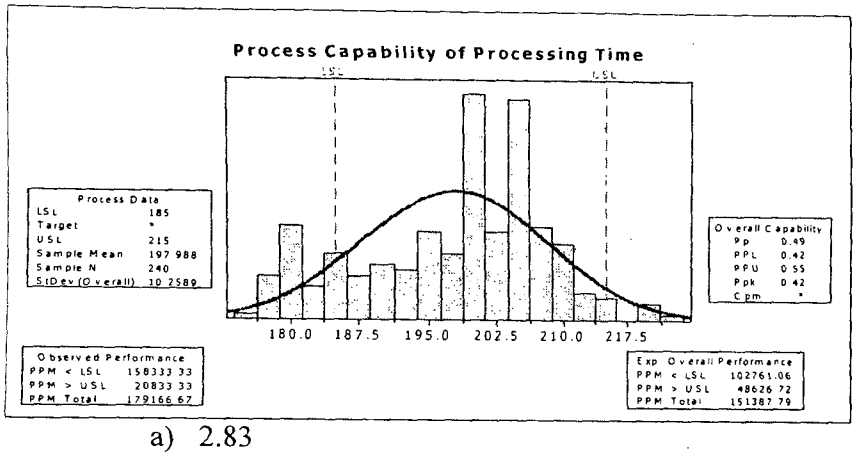
24. Match the following:

A-validation of the root causes	1. Define phase
B-SIPOC	2. Measure phase
C-Ranking the solutions	3. Analyze phase
D-Process capability evaluation	4. Improve phase

a) A3-B1-C2-D4
b) A3-B1-C4-D2
c) A4-B2-C3-D1
d) A2-B1-C3-D4

[5]

25. The approximate Sigma Level of the process is



a) 2.83
b) 2.42
c) 1.58
d) 1.42

[5]

26. The COPQ is defined as the following:

a) External failure costs
b) Internal and external failure costs
c) Appraisal and failure costs
d) Prevention, appraisal and failure costs

[2]

27. For the following project prioritization matrix, which projects need to be selected?

	Project Assessment Criteria					
	Budget Requirement	Savings Potential	Customer Benefits	Project Duration	Team Quality	Data Quality
Weight→	10%	15%	20%	20%	15%	20%
Project 1	26%	10%	18%	14%	20%	6%
Project 2	26%	30%	18%	14%	20%	6%
Project 3	26%	30%	18%	4%	20%	56%
Project 4	9%	10%	10%	14%	20%	6%
Project 5	9%	10%	18%	14%	10%	20%
Project 6	4%	10%	18%	40%	10%	6%

a) 1, 2, 3
b) 1, 4, 5
c) 3, 4, 5
d) 2, 3, 6

[5]

28. The three key elements to statistical thinking are:

a) Process, variation and data
b) Process, variation and statistical tools
c) Variation, data and statistical tools
d) Variation, statistical tools and statistical inferencing

[2]

29. From the strategic point of view, the balanced scorecard should consider the following parameters:

a) Price, quality, service, & availability
b) Price, quality, service, & brand
c) Learning & growth, internal processes, customer, & financial
d) Quality, brand, cost efficiency, & revenue growth

[2]

30. QFD is

a) A systematic way to document and break down the suppliers' needs
b) A structured methodology to identify and translate customer needs into technical requirements
c) A multidimensional matrix that shows the correlation between "what's" and "when's"
d) All of the above

[2]

31. You have been asked to sample a lot of 500 units from a vendor whose past quality has been about 2% defective. A sample of 40 pieces is drawn from the lot and you have been told to reject the lot if you find two or more parts defective. What is the probability of finding 2 or more parts defective?

a) 0.953
b) 0.809
c) 0.191
d) 0.047

[5]

32. The new seven Japanese management and planning tools are:

a) Affinity diagram, Tree diagram, PDPC, Matrix diagram, Gantt chart, Prioritization matrix, Arrow diagram
b) Affinity diagram, Tree diagram, PUGH analysis, Matrix diagram, Gantt chart, Prioritization matrix, Arrow diagram
c) Affinity diagram, Tree diagram, PUGH analysis, GEMBA analysis, Gantt chart, Prioritization matrix, Arrow diagram
d) Affinity diagram, Tree diagram, PDPC, Matrix diagram, Interrelationship diagram, Prioritization matrix, Arrow diagram

[2]

33. When a process is documented and validated, one can analyze it for some of the following specific problem areas:

a) Strategic activities, Disconnects, Bottlenecks, Redundancies, Decisions/Inspections
b) Disconnects, Bottlenecks, Redundancies, Rework loops, Decisions/Inspections
c) Disconnects, Bottlenecks, Strategic activities, Tactical activities, Decisions/Inspections
d) Strategic activities, Tactical activities, Operational activities, Bottlenecks, Decisions/Inspections

[2]

34. A Belt is attempting to improve the soldering on a micro-processor used for a new hand-held device. As a result he should build a list of Critical to Quality Characteristics based on —

a) Service benefits
b) Product features
c) Price
d) Size of unit

[2]

35. Producing more than is needed by the next step in the process or more than the customer needs is an example of which of the Seven Elements of Waste?

a) Overproduction
b) Correction (defects)
c) Inventory
d) Motion

[2]

INDIAN STATISTICAL INSTITUTE
Semester Examination: 2016-17
Course Name: M.Tech (QR & OR), 2nd YEAR (E & S Streams)

Subject Name: Six Sigma

Date: 05/12/2016 **Maximum Marks: 100**

Duration: 3 hrs.

Answer All Questions in the Answer Sheet. Work out Where It is required for Your Answer.

1. In Six Sigma, the Measure Phase entails:

   a) Using SPC to ascertain the repeatability and reproducibility of metrics in an operational environment and setting up control plans for tolerances, controls, measures and standard operating procedures
   b) Validating gaps in requirements versus current metrics with vital causes, quantifying opportunity to close gaps and prioritizing root causes and identifying the most contributing one
   c) Selecting product characteristics, mapping the respective processes, making the necessary measurement, recording the results, and estimating the short – and long-term process capabilities
   d) Preparing a clearly defined problem statement and possible symptoms identified along with a goal statement defining the results we are seeking with a measurable target

   [2]

2. One of the main objectives of Measure Phase is

   a) To use SPC for validating collection systems and to document the follow-up steps
   b) To establish standards for performance that are based on actual customer input, so that process effectiveness and capability can be accurately measured
   c) To weigh the costs and benefits of "quick-hit" versus more difficult solution options, prioritize solution options for each root cause and examine solutions with a short-term and long-term approach
   d) To identify the most contributing KPIVs and KPOVs through statistical analysis and test the significance of them

   [2]

3. Process capability is a ratio defined as

   a) $\frac{\text{VOP}}{\text{VOC}}$
   b) $\frac{\text{VOC}}{\text{VOP}}$
   c) $\frac{\text{PRECISION}}{\text{TOLERANCE}}$
   d) $\frac{\text{Variability due to gage}}{\text{Product Variability}}$

   [2]

4. The characteristics of a VOC metrics are as follows:

a) Performance, reliability, durability, serviceability, aesthetics, features
b) Tangibles, reliability, responsiveness, assurance, empathy
c) Credibility, reliability, precision, accuracy, action ability, predictability
d) System / process, commitment, competence, communication, continuous improvement

[2]

5. The gage repeatability is

a) The variability due to different operators using the gage
b) The difference between the gage measurement value and the "true" value of the measured characteristic
c) The basic inherent precision of the gage itself
d) The measure of the change in accuracy through the operating range of the gage

[2]

6. The gage reproducibility is

a) The variation associated with repeated measurements of a single part by one operator using the gage
b) The demonstration that no special causes of variation are present
c) The basic inherent precision of the gage itself
d) The between-operator variation associated with repeated measurements of a single part by different operators using the gage

[2]

7. Which of these defects per million opportunity (DPMO) levels approximate to four sigma quality?

a) 308,537
b) 66,807
c) 6,210
d) 233

[2]

8. Which of these is the best definition of "Takt Time"?

a) Pace of a manufacturing system as determined by customer demand
b) Average time taken to complete a task under normal operating condition
c) The maximum operating speed of a process or production line
d) The long-term historic performance capacity of a system

[2]

9. In an attribute agreement analysis, to estimate the Kappa value one needs to compute

a) $$\frac{\text{Total observation - No. of expected agreement}}{\text{No. of observed agreement - No. of expected agreement}}$$

b) $$\frac{\text{1 - Expected proportion by chance only}}{\text{Observed proportion - Expected proportion by chance only}}$$

c) $$\frac{\text{Observed proportion - Expected proportion by chance only}}{\text{1 - Expected proportion by chance only}}$$

d) $$\frac{\text{No. of expected agreement - No. of observed agreement}}{\text{Total observation - No. of expected agreement}}$$

[2]

10. A cake factory has received an order to ship 1000 cakes in 30 days. The factory operates for two shifts in a day and for 26 days in a month. The duration of a shift is 8 hours. While for the day shift there is a 30 minutes lunch break, 2 breaks of 15 minutes duration each, 10 minutes for a team briefing and 10 minutes for a basic maintenance checks; for the afternoon shift except lunch break other breaks remain intact. The takt time of the process is:

a) 20.02 minute
b) 21.58 minute
c) 24.58 minute
d) 24.90 minute

[5]

11. The following table contains the data needed for the OEE computation.

Parameter	Data
Shift length	8 hours
Short break	1 @ 10 minutes
Lunch break	1 @ 30 minutes
Downtime	45 minutes
Ideal run rate	70 pieces/minute
Total pieces	20,000 pieces
Rejected pieces	500 pieces

The pertinent OEE approximately equals to

a) 73%
b) 65%
c) 63%
d) 67%

[5]

12. Kappa values range from

a) $-\infty$ to $+\infty$
b) 0 to +1
c) 0 to $+\infty$
d) −1 to +1

[2]

13. The length of three parts are additive in an assembly. Their design specifications for length and tolerance are 0.240 ± 0.006, 0.3200 ± 0.0006, and 1.360 ± 0.003, respectively. Assume that each of the distributions is normal. Combine these dimensions statistically to give a final length and tolerance to three decimal places.

a) 1.360 ± 0.006
b) 1.920 ± 0.565
c) 1.920 ± 0.009
d) 1.920 ± 0.007

[2]

14. If we want to use a scale where the relative order of the measurements and also the difference between them are important, then we should use

a) The ordinal scale of measurement

b) The interval scale of measurement
c) The nominal scale of measurement
d) The ratio scale of measurement

[2]

15. The standard deviation for a part with underlying normal distribution is $\sigma = 0.25$. How large a sample must we take to be 95% confident of getting a margin of error of 0.025 of the true mean?

a) 100
b) 385
c) 96
d) 358

[2]

16. If two sigma limits are substituted for conventional three sigma limits on a control chart, one of the following occurs

a) Decrease in α risk
b) Increase in β risk
c) Increase in α risk
d) No change in α or β risk

[2]

17. A manufacturer of computer hard drives wants to measure their Sigma Level. Over a given period of time, the manufacturer creates 83,934 hard drives. The manufacturer performs 8 individual checks to test quality of the drives. During testing 3432 are rejected. The short-term sigma level is

a) 1.75
b) 4.07
c) 2.57
d) 3.25

[5]

18. If mean = 10, median = 8 and standard deviation = 6, then the coefficient of skewness is

a) 1
b) -1
c) 2/6
d) 2

[2]

19. The Pearson kurtosis is defined as:

$$Ku = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s}\right)^4$$

choose the right alternative about its values

a) Leptokurtic = 3; Platykurtic < 3; Mesokurtic > 3
b) Mesokurtic = 3; Platykurtic > 3; Leptokurtic < 3
c) Leptokurtic > 3; Platykurtic < 3; Mesokurtic = 3
d) Mesokurtic < 3; Platykurtic > 3; Leptokurtic = 3

[2]

20. An automobile manufacturer receives steering subassemblies from three vendors – 40% from vendor 1, 35% from vendor 2, and 25% from vendor 3. Quality performances for the three vendors differ. The percentage of defective subassemblies from vendors 1, 2, and 3 are 2, 2.6,

and 3%, respectively. If a randomly picked subassembly is found to be defective, the probability that it is from vendor 1 is

a) 0.0246
b) 0.0260
c) 0.3077
d) 0.3252

[5]

21. The difference in the proportion of nonconforming components produced by two processes is to be established. A random sample of 96 parts taken from one process produced 6 nonconforming parts and a sample of 120 parts taken from the second process had 8 nonconforming parts. The 95% confidence interval for the difference in proportion of nonconforming parts is

a) $-0.0701 \le p_1 - p_2 \le 0.0617$
b) $-0.0667 \le p_1 - p_2 \le 0.0817$
c) $-0.0887 \le p_1 - p_2 \le 0.0617$
d) $-0.0777 \le p_1 - p_2 \le 0.0817$

[5]

22. Cylindrical pins are produced on a centerless grinding machine. A random sample of 10 cylindrical pins has been collected and the standard deviation of the sample is obtained as 0.025 mm (25 μm). The 95% confidence interval for the population standard deviation (σ) is:

a) $0.0296 \le \sigma \le 0.0208$
b) $0.0179 \le \sigma \le 0.0356$
c) $0.0172 \le \sigma \le 0.0456$
d) $0.0003 \le \sigma \le 0.0021$

[5]

23. For signal-response systems where the relation signal-response is linear, the indicator of robustness that needs to be maximized is defined as follows:

a) $SN = 10\log\left(\frac{\bar{y}^2}{s^2}\right)$
b) $SN = -10\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right)$
c) $SN = -10\log\left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{y_i^2}\right)$
d) $SN = 10\log\left(\frac{\beta^2}{s^2}\right)$

[2]

24. The Headquarters of a bank assigns a Six Sigma team to investigate the deposit process and reduce the number of complaints at the branches. The bank provides the practitioners with 380 deposit slips (units). A collection of data after each step revealed the following:

- Customer gives deposit slip, cash and checks to bank teller (5 defects)
- Bank teller (associate) processes customer's deposits (6 defects)

- Bank Associate enters deposit in Bank's electronic system (3 defects)
- System generates deposit confirmation document (4 defects)
- Customer receives deposit confirmation document from Bank Associate (2 defects)

The benchmark sigma level of the deposit process is:

a) 2.19
b) 3.69
c) 2.96
d) 3.81

[5]

25. The management at a large auto manufacturing company has drafted a new labor management contract. Before this contract is presented to and negotiated with labor union, the management wants to seek opinion of workers on this proposed contract. The management selected a random sample of 300 workers and asked their opinions. The responses of these 300 workers are presented in the table below.

Gender	Support	Against	No Opinion
Female	87	32	6
Male	93	70	12

Using the 1% significance level, what can you conclude about the gender and opinions of workers?

a) Gender and opinions are independent
b) Gender and opinions are dependent
c) Cannot be concluded since data are insufficient
d) The level of significance is non-conventional; so no conclusion can be made

[5]

26. The objective of failure mode and effects analysis is to

a) Anticipate product failures and prevent them from occurring
b) Devise ways of minimizing the impacts of product failures when they occur
c) Describe the interrelationships among product failures
d) Quantify the likelihoods of different product failures

[2]

27. Assume that the cost of repairing a failed television set in the factory is Rs. 2000 per unit. Compare the losses from the options below caused by deviations from the target value for two television sets, one produced in Factory A with distribution of color concentration as Normal and the other produced in Factory B with Uniform distribution of color concentration. The tolerance interval ranges from $T - \Delta$ to $T + \Delta$, where $\Delta = 5$ for the target T.

a) Loss in Factory A = Loss in Factory B
b) Loss in Factory A = 3 * Loss in Factory B
c) Loss in Factory B = 3 * Loss in Factory A
d) Loss in Factory B = 2 * Loss in Factory A

[5]

28. In Taguchi's β – correction for process control, the optimum value of β is given by

a) $\beta = 1 - \frac{1}{F}$ *where* $F = \frac{(\hat{\mu} - \mu_0)^2}{\sigma^2}$ $\beta = 0$ *if* $F \leq 1$

b) $\beta = 1 + \frac{1}{F}$ *where* $F = \frac{(\hat{\mu} - \mu_0)^2}{\sigma^2}$ $\beta = 0$ *if* $F \leq 1$

c) $\beta = \frac{1}{F} - 1$ *where* $F = \frac{(\hat{\mu} - \mu_0)^2}{\sigma^2}$ $\beta = 0$ *if* $F \leq 1$

d) $\beta = 1 - \frac{1}{F}$ *where* $F = \frac{(\hat{\mu} - \mu_0)^2}{\sigma^2}$ $\beta = 0$ *if* $F < 1$

[2]

29. Juran had described the 'concept of dominance' and the category of dominance of a process as follows:

a) Set up, time, component, worker, method
b) Set up, time, information, worker, method
c) Set up, time, component worker, material
d) Set up, time, component, worker, information

[2]

30. Some of the types of errors dealt with the Poke Yoke system are as follows:

a) Machine Readability, Lack of Standards, Lack of Experience
b) Identification, Misunderstanding, Forgetfulness
c) Neither a nor b
d) Both a and b

[2]

31. The results of a 2-factor experiment with factor A at 4 levels and factor B at 5 levels are shown below.

Factor A	Factor B				
	B_1	B_2	B_3	B_4	B_5
A_1	76	67	81	56	51
A_2	82	69	96	59	70
A_3	68	59	67	54	42
A_4	63	56	64	58	37

At a 1% level of significance, which of the following is the finding of the analysis?

a) A is significant but B is not significant
b) B is significant but A is not significant
c) Neither A nor B is significant
d) Both A and B are significant

[5]

32. The 5 "W"s to be considered during a Root Cause Analysis are:

a) Who, What, When, Where and Why
b) Who, Who, Who, Who and Who
c) Why, Why, Why, Why and Why
d) Why, When, Why, When and Why

[2]

33. Value Stream Mapping exercise begins with:

a) Customer specifications
b) A consensus about how the process should be executed

c) A map of the current status (process and information flow)
d) Best practices implementation

[2]

34. Which one is not a goal of the Define phase?

a) Identify the voice of the customer
b) Define a data collection plan
c) Develop high level process map
d) Create project plan and milestones

[2]

35. What is the main goal in the Measure phase?

a) Define precisely the project charter
b) Understand the problem
c) Identify the root causes
d) Implement actions
e) Ensure results are sustainable

[2]

INDIAN STATISTICAL INSTITUTE

Semestral Examination : 2016-17

M. TECH. (QR & OR) II YEAR

Applied Stochastic Processes and Time Series Modelling

Date: 05 December 2016 Maximum Marks: 100 Duration: 3 hours

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

1. Fill in the blanks with justification wherever asked for.

   (a) Suppose that $\{N(t), t \geq 0\}$ is a Poisson process. Then

   $$s < T_n \leq t \quad \text{if and only if} \quad N(s)\_\_\_\_\_,$$

   where T_n is the time of occurrence of the n-th event.

   (b) Let $X_1, X_2, \ldots$ be the inter-occurrence times of the events of a nonhomogeneous Poisson process. Then the random variables X_i's are _____ distributed.

   (c) Suppose that $\{N(t), t \geq 0\}$ is a renewal process induced by $\{X_n, n \geq 1\}$. Then $\mathrm{E}[N(t)|X_1 = x]$ equals _____. (Justify.)

   (d) Let $\{N(t), t \geq 0\}$ be a renewal process with mean inter-occurrence time as μ. Then $\mathrm{E}[S_{N(t)+1}]$, in terms of μ, is equal to _____. (Justify.)

   [2+2+3+3 = 10]

2. The time between successive renewals, for the renewal process $\{N(t), t \geq 0\}$, is a continuous random variable whose probability density function is the following:

   $$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 < x < \frac{1}{2} \\ \\ \frac{3}{2} & \text{if } \frac{1}{2} \leq x < 1. \end{cases}$$

   (a) Show that $m(t) = \exp(t/2) - 1$ for $0 < t < 1/2$.

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

   [5+8 = 13]

3. Let $\{N(t), t \geq 0\}$ be a renewal process induced by $\{X_n, n \geq 1\}$ with $\mu = \mathrm{E}[X_1]$. Prove that, with probability 1,

   $$\frac{N(t)}{t} \longrightarrow \frac{1}{\mu} \quad \text{as} \quad t \to \infty.$$

   [10]

4. Processing times of certain jobs by the two machines M_1 and M_2 have Gamma$(4, 2)$ and U$(0, 4)$ distributions respectively. Approximate the probability that the two machines together complete at least 90 jobs by time $t = 100$. **[12]**

[P.T.O.]

5. The two machines, in a shop, operate with the same failure rate μ, and there is a repair facil which can repair one machine at a time with rate λ. Let $X(t)$ be the number of operat nal machine(s) at time t. Assume $\{X(t), t \geq 0\}$ to be CTMC having stationary transiti probabilities.

   (a) Write down the state-space and the generator.

   (b) Obtain stationary distribution of the process.

[(2+3)+5 = 1

6. (a) Consider the time series process $X_t = 4 + 0.4X_{t-1} + 0.5X_{t-2} + Z_t$, $Z_t \sim \text{WN}(0, \sigma^2)$. (a) Verify whether $\{X_t\}$ is a stationary. (b) Find the mean of the process.

   (b) What is variogram? Explain the behaviour of variogram for a stationary process.

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

7. (a) Let $\{Z_t\}$ be a stationary process with mean zero and variance σ^2. If $X_t = (a + bt)s_t + Z$ where s_t is a seasonal component with period 12 such that $s_t = s_{t-12}$, and a and b a constants. Show that $W_t = (1 - B^{12})^2 X_t$ is a stationary process.

   (b) Show that one-step-ahead forecasting for an ARIMA(0, 1, 1) process is simple exponenti smoothing forecast.

[6+6 = 1

8. Let $\{X_t\}$ be a sequence of random variables with mean 0 and unit variance. Define $Z_t = X_t X_{t-}$ Examine whether $\{Z_t\}$ is stationary or not for the following cases: (a) $\{X_t\}$ is an i.i.d. proce and (b) $\{X_t\}$ is a white noise process. **[4+3 =**

9. Consider the following time series data on weekly sales (in 1000s gallon) of gasoline.

Week	1	2	3	4	5	6	7	8	9	10	11	12
Sales	17	21	19	23	18	16	20	18	22	20	15	22

   (a) Calculate the forecast value of week 13 at week 12 by exponential smoothing metho taking smooting parameter $\alpha = 0.2$. Take $\hat{x}_1(1) = \bar{x}$, the sample mean.

   (b) Suppose an MA(1) process

$$X_t = \mu + Z_t - \theta Z_{t-1}, \text{ where } |\theta| < 1 \text{ and } Z_t \sim \text{WN}(0, \sigma^2)$$

   is to be fitted for the above data. Calculate the moment estimates of μ, θ and σ^2. Forecas the sales for weeks 13 and 14 at week 12 using the fitted model. Take $\hat{x}_1(1) = x_1$, the fir observation.

[8 + (10 + 8) = 2

*** xXx ***

INDIAN STATISTICAL INSTITUTE

Semester Examination: 2016-17

(First Semester)

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

Subject Name : Reliability II

Date: 07/12/2016 Maximum Marks: 100 Duration: 3 hours

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

1. a) Define increasing failure rate average (IFRA) distribution and decreasing failure rate average (DFRA) distribution in the context of life distribution.

   b) Show that F is IFRA if and only if $\bar{F}(\alpha t) \geq \bar{F}^{\alpha}(t)$, for all $0<\alpha<1$ and $t \geq 0$

   c) Prove that IFRA distributions possess closure property.

   [2+4+14=20]

2. a) Describe the problem of optimizing reliability allocation to each of the subsystems of a given system using effort minimization algorithm.

   b) Describe specifically the assumptions to be made on the effort function for transforming the reliability level of the subsystem from R_i to R_i^* $(R_i^* > R_i)$.

   c) State the solution to the above problem (proof not required).

   d) A system consists of six subsystems that must function if the system has to function properly. The system reliability goal is 0.998. The present reliability of the subsystems are 0.70, 0.75, 0.80, 0.85,0.90 and 0.95 calculated for 50 hours, The system mission time is 20 hours. Use ARINC methodology for determining target reliability to be achieved by the subsystems in order to achieve a system reliability target of 0.998 and find out target reliabilities for the subsystems.

   [5+4+2+14 = 25]

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

   [10]

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

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

[5+15=20]

Q5. 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.

[7+4+14 = 25]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination: 2016-17

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

Subject Name: Advanced Multivariate Analysis

Date: 09 .12.2016 Maximum Marks: 100 Duration: 3½ hours

Note: This paper carries 120 marks.
You can answer any part of any question, but maximum you can score is 100.
All notations 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) Principal components do not always lead to meaningful interpretation.

iii) One can go back to the original variables in the case of principal component regression.

iv) Discriminant function analysis is a dependence technique .

v) A model developed by multiple linear regression method represents the underlying causal model.

vi) 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 $y = \beta_0 x_1^{\beta_1} x_2^{\beta_2}$.

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

viii) Factor analysis is a dependence technique.

ix) MANOVA is an interdependence technique.

x) Principal Component Analysis is an interdependence technique.

[2 x 10 =20]

2) (a) What is multicollinearity?

(b) How is multicollinearity diagnosed?

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

(d) Discuss the Ridge Regression or Principal Component Regression method for overcoming the problem of multicollinearity.

[3 + 3 + 4 + 10 = 20]

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

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

c) Define communality and specificity.

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

e) Explain how orthogonal transformation help in factor extraction.

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

[6+ 4 + 4 + 8 + 4 + 4 = 30]

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 are the differences between classification analysis and cluster analysis?

(b) What are the differences between hierarchical and partitioning method of clustering?

(c) If we are to cluster objects based on observations on a set of variables what is the suitable measure of dissimilarity?

(d) Consider the following set of bivariate data. Cluster the objects using an appropriate agglomerative clustering technique (construct the distance matrix by Euclidean distance measure). Decide on the number of clusters.

Objects	Observations	
	X_1	X_2
A	1	1
B	1	2
C	6	3
D	8	2
E	8	4

[4 + 3 + 1 + (9 + 3) = 20]

INDIAN STATISTICAL INSTITUTE

First Semestral Examination : 2016-17

M. Tech. (QR & OR)-II

Industrial Experimentation

Date: 13/12/2016 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. What is meant by randomization? What role does it have in industrial experimentation? Do you think that the "experiments are usually iterative" and why?

[2+6+5 = 13]

2. What is a Latin square design? How to apply the principle of randomization to such a design? What are the ways of replicating a Latin square design?

[3+5+6 = 14]

3. (a) What is a factorial experiment? What are its advantages over a one-at-a-time experiment?

(b) The sum of two replicates for each of the 16 treatment combinations $(1), a, b, ab, c, ac, bc, abc, d, ad, bd, abd, cd, acd, bcd, abcd$ are 4, 7, 5, 6, 7, 6, 8, 6, 4, 10, 5, 10, 8, 9, 7, 10 respectively. Estimate the factor and interaction effects. Write down a regression model for predicting yield, assuming that (i) all the four factors may vary over the range from -1 to +1 (in coded units), (ii) three factor and higher interactions are negligible, and (iii) any effect, lower than 4.0 in magnitude, is insignificant.

[(3+6)+(12+4) = 25]

4. Discuss the procedure for confounding a 2^5 full factorial design in four blocks of size eight each with an example. Obtain the treatments in the principal block of such a design and what are the effects confounded with the blocks in your design? How can one obtain the principal block when a block other than the principal block is given?

[6+5+3+2 = 16]

5. List the treatment combinations of a block, other than the principal block for a 3^{4-2} fractional factorial design having the generating relations $I = AB^2C$ and $I = BCD$. What is the defining relation for your design? What is its resolution? Write the alias structure for effect D.

[8+2+1+4 = 15]

6. The factors that influence the breaking strength of a synthetic fibre are being studied. Three polymer type (A), four production machines (B) and three operators (C) are chosen and a factorial experiment is run. The polymer type is a fixed factor, and production machines and operators are random. There are two replications. The linear model, with usual assumptions, used for the design is:

$$y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \epsilon_{ijkl}$$

where τ_i is the effect of factor A, β_j is the factor B effect and γ_k is the effect of the factor C.

The analysis of variance table for the breaking strength data is given in Table 1.

Table 1: ANOVA for the Breaking Strength Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	EMS	$F_{observed}$	Remarks
A		1023.36				
B		423.82				
C		7.19				
AB		1211.97				
AC		137.89				
BC		209.47				
ABC		166.11				
Error						
Total		3950.32				

(a) Complete the ANOVA table, based on general rules, up to the column EMS starting from the column degrees of freedom.

(b) Test for the significance of main effect of A and interaction effect of A and B.

[(2+1+2+10)+(5+2) = 22]

7. To reduce the production time of a certain job the effects of seven factors on job processing time are to be studied. The factors and their levels are given in Table 2.

Table 2 : Factors and their levels

Factor code	Factor	Levels			
		I	II	III	IV
A	Machine used	1	2	-	-
B	Operator *	1	2	3	4
C	Power setting	1	2	-	-
D	Type of tool used	1	2	-	-
E	Tool angle	1	2	-	-
F	Spin speed	1	2	-	-
G	Duration of spin	1	2	-	-

* : For practical reasons, different groups of four **operators** are to be used on different machines

Design an experiment in sixteen runs, using the *OA* table and Taguchi's associated "linear graph", to study all the main effects along with interactions *AC*, *AF* and *AG*. Give the experimental layout showing the levels of factors *A*, *B* and whatever factor is assigned to one of the columns 8 or 9, and for others only indicate the column assignments.

[15]

----- X -----

TABLE 3: Orthogonal Array - OA(16,15,2,2) or $L_{16}(2^{15})$

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

TABLE 4: Interaction between columns

L_{16} (2^{15})

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

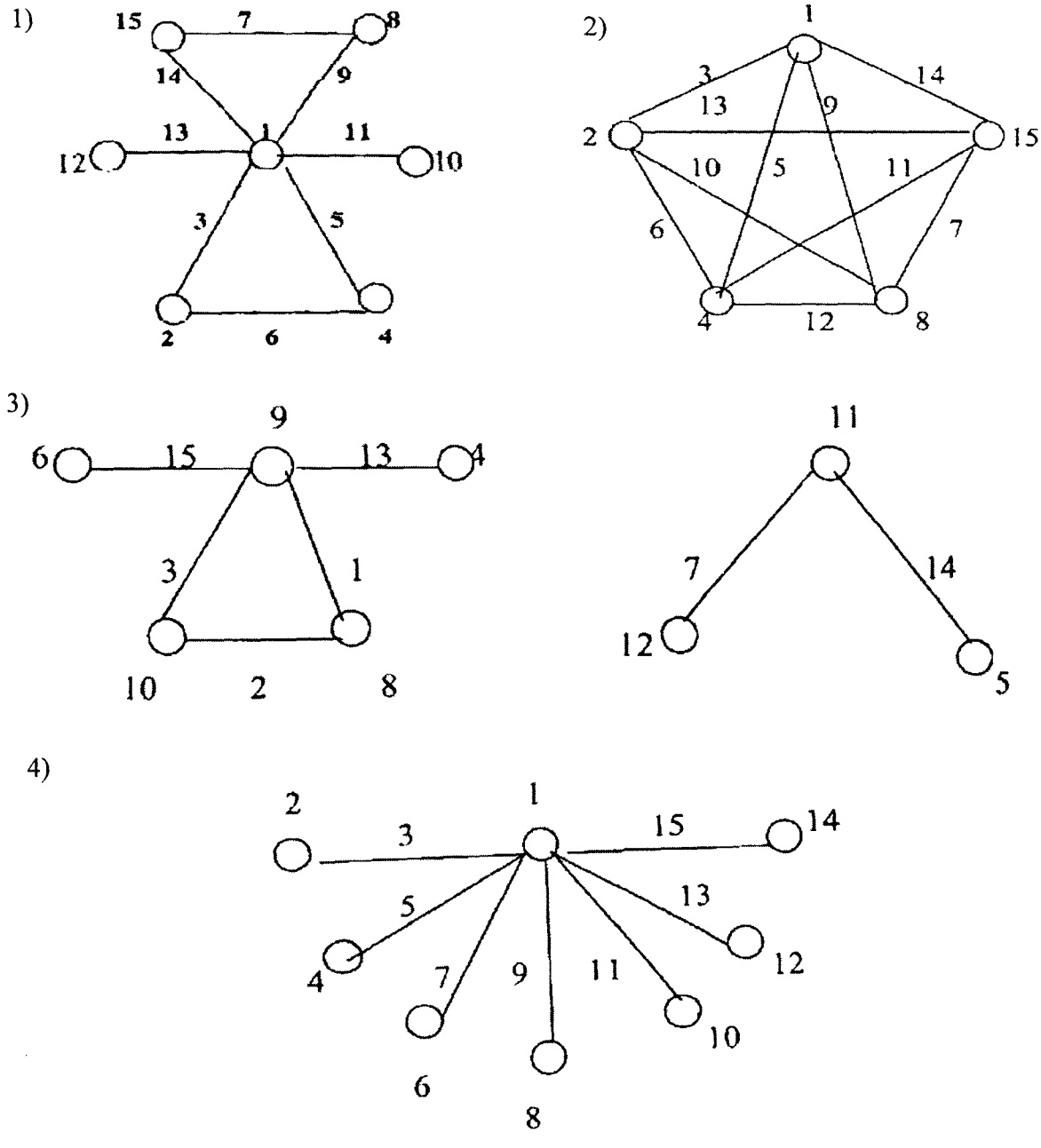


Fig. 1: Some Linear Graphs of OA(16,15,2,2)

Table 5: F distribution (5%) Table

$F_{0.05, v1, v2}$

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

INDIAN STATISTICAL INSTITUTE

FIRST SEMESTER EXAMINATION: 2016 – 17

Course Name: M Tech (QROR) 2nd Year

Subject Name: Business Analytics

Date: 14.12.2016 **Maximum Marks: 100** **Duration: 3 hours**

Notes, if any: Answer question 6 and any 3 from the rest

1. A bank is carrying out fraud analytics. They are trying to develop a model to find out whether a particular credit or debit card transaction is fraudulent or not.
   a. Is it a supervised or unsupervised analytics? Why? [3]
   b. Would you classify this primarily as a predictive analytics or an explanatory analytics? [2]
   c. Suggest two possible techniques that you may use to solve this problem. Explain the techniques briefly giving justification. [6 + 6 = 12]
   d. Explain briefly the concepts of test data and training data. [2 + 2 = 4]
   e. Explain how you will understand whether your model has been over fitted or not. [4]

2. Answer the following
   a. What is a classification problem? [2]
   b. Suppose you have fitted a binary logistic regression to assess whether customers applying for loan to a bank are likely to return the loan or not. Suppose the model identifies non defaulters with 92% accuracy.
      i. You have noted that the overall misclassification is only 15%. Can you say that the model is "good" as far as classifications are concerned? [4]
      ii. What would be your interpretation if the area under the ROC curve was 0.82? Explain the concept behind the area under ROC curves? [3 + 5 = 8]
      iii. Explain the concept of Linear Discriminant Analysis as a classification model. [7]
      iv. In what situation would LDA be a better classifier than logistics regression? When would the logistic regression be better? Explain briefly. [2 + 2 = 4]

3. Suppose while fitting a multiple linear regression model, an analyst computed the VIF for model validation and found that its value is 8.6. The analyst decided to use Ridge regression to take care of the situation
   a. What is ridge regression? Explain briefly and write the model clearly. [4]
   b. Explain how the regression coefficients and the shrinkage parameters are estimated for a ridge regression. [6 + 3 = 9]
   c. Do you think application of ridge regression will take care of the problem? Why? [2]
   d. When will a ridge regression be the same as a usual regression model? Explain. [2]
   e. What transformations need to be carried out on the input data while fitting a ridge regression model? Why? [3]
   f. What is high dimensional data? Can you use best subset method to take care of high dimensional data? [2 + 3 = 5]

P.T.O

4. An automobile company is trying to fit a model to estimate the fuel efficiency of its cars for various speed ranges. The concerned engineers noted that the form of relationship between fuel efficiency and speed various across different ranges of speed. For example, in some range, relationship between fuel efficiency and speed may be linear and in some other range it may be quadratic. The engineers identified some speed ranges and fitted different regression models for the different ranges without taking any other constraint into consideration
    a. Do you envisage some problem with the approach adopted by the engineers? Why? [3]
    b. Explain the concept of regression splines in this context. How does regression spline take care of the problem stated above? Write the model. [5]
    c. Explain how the truncated power functions are used to fit cubic splines. In this context explain how many knots are typically chosen and what methods may be followed to decide about the number of knots used. [7 + 5 = 12]
    d. Explain the concept of natural splines. [5]

5. In supervised analytics we need to fit a model of the form $y = f(x)$ – where x gives a vector of input variables. We know that in practice, we may need to fit a very flexible model
    a. Explain what is meant by a flexible model. [2]
    b. Write the Projection Pursuit Regression (PPR) model. Is it a flexible model? [4 + 2 = 6]
    c. Write the neural network model with one hidden layer and show that this model is essentially same as PPR model. [6 + 3 = 9]
    d. Explain how the parameters of the neural network model for classification with one hidden layer may be estimated using the technique of maximum likelihood. Write the models clearly. [8]

6. Give brief answers to the following
    a. What are Leave One Out Cross Validation and K – Fold Cross Validation? [5]
    b. What is the difference between bagged trees and random forest? [5]
    c. Write the value estimation model for K nearest neighbours? What happens when the value of K is 1? [4 + 1 = 5]
    d. Explain how Gini Index and Cross Entropy are used to measure node purity for a classification tree. [5]
    e. Explain the concept of basis functions in the context of regression. Write a step function model using indicators functions as the basis functions. [5]

INDIAN STATISTICAL INSTITUTE

First Semester EXAMINATION: 2016

Course Name: M. TECH (QROR) II

Subject Name: OR II

Date: 16.12.2016 **Maximum Marks: 100** **Duration: 3 hours**

Answer 1. and any four from 2. To 7.

1. a) Formulate a quadratic programming problem as linear complementarity problem LCP (q, M). State the similar formulation in case of linear programming problem.

   b) Consider an LCP (q, M) where

$$M = \begin{bmatrix} 4 & 0 & 2 & 1 \\ 1 & 5 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 \end{bmatrix} \qquad q = \begin{bmatrix} 2 \\ 2 \\ 1 \\ -4 \end{bmatrix}$$

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

[12+8=20]

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

   b) State the KKT sufficient conditions of optimality.

[6+9 = 15]

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

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

[6 + 9 = 15]

P.T.O

INDIAN STATISTICAL INSTITUTE
M. Tech. (QR & OR) 2nd YEAR
Year: 2016
MID SEMESTER SUPPLIMENTARY EXAMINATION

Subject: Operations Research-II

Date of Exam: 19.12.2016 **Max. Marks: 100** **Time: 3 hours**

Answer any Five.

1. Define the following concepts of game theory:
   i. Pure and Mixed strategy.
   ii. Pay-off matrix.
   iii. Two person zero-sum game.
   iv. Equilibrium or Saddle point.
   v. Solve the game whose pay-off matrix is given below:

$$\begin{bmatrix} 30 & 40 & -80 \\ 0 & 15 & -20 \\ 90 & 20 & 50 \end{bmatrix}$$

[3+3+3+3+8=20]

2. Describe the branch and bound algorithm to solve an integer programming problem. State the method to formulate an integer programming problem into binary integer programming.

[15+5=20]

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

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

[20]

4. Formulate the following game problem as linear programming problem and solve using simplex algorithm:

$$\begin{bmatrix} 1 & -1 & 3 \\ 3 & 5 & -3 \\ 6 & 2 & -2 \end{bmatrix}$$

[20]

5. What is flows in network? Given an example to state the max-flow min-cut problem. State Ford Fulkerson method.

[6+6+8=20]

6. Describe all integer 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?

[10+5+5=20]

INDIAN STATISTICAL INSTITUTE
Mid-Semestral (Supplementary) Examination : 2016-17
M. TECH. (QR & OR) II YEAR
Applied Stochastic Processes and Time Series Modelling

Date: 20 December 2016 | Full Marks: 50 | Duration: 2 hours

Notation have usual meaning.

1. Suppose that $\{N_1(t), t \geq 0\}$ and $\{N_2(t), t \geq 0\}$ be two independent Poisson processes with rates λ_1 and λ_2 respectively, and $N(t) = N_1(t) + N_2(t)$ for all $t \geq 0$. Find the probability that the first event of the combined process $\{N(t), t \geq 0\}$ comes from $\{N_1(t), t \geq 0\}$.

[10]

2. The failures of certain device occur according to non-homogeneous Poisson process whose intensity function $\lambda(t)$ is given by:

$$\lambda(t) = \begin{cases} 0.2 & \text{if } 0 \leq t \leq 10 \\ 0.3 & \text{if } t > 10, \end{cases}$$

where t is the age (in years) of the device.

(a) Calculate the probability that a five-year-old device (without failure) will have exactly two failures over the next 10 years.

(b) Knowing that the device had exactly one failure in course of the first 5 years of the 10 years considered in (a), what is the probability that this failure took place during its sixth year of use?

[5+7 = 12]

3. The number of accidents in a town is in accordance with a Poisson process having rate 2 per day, and the number X_k of people involved in the k-th accident has the distribution:

$$\mathrm{P}[X_k = i] = \frac{1}{2^i}, \quad i = 1, 2, 3, \ldots,$$

for $k = 1, 2, \ldots$. Let $Y(t)$ be the number of people involved in accidents during $[0, t]$ days. Assume that X_k's are independent among themselves, and are independent of the Poisson process.

(a) Derive the moment generating function of $Y(t)$.

(b) Hence, find the mean and variance of the number of people involved in accidents during a week.

[7+6 = 13]

[P.T.O.]

4. (a) Define a strong stationary process.

   (b) Let $\{X_t\}$ be a process given by

$$X_t - \phi X_{t-1} = Z_t, \quad |\phi| < 1 \text{ and } \{Z_t\} \sim \text{WN}(0, \sigma_z^2).$$

   Define a process $\{Y_t\}$ such that $Y_t = X_t + W_t$, where $\{W_t\} \sim \text{WN}(0, \sigma_w^2)$ and $E(W_s Z_t) = 0$ for all s and t. (i) Find lag-1 autocorrelation of the process $U_t = Y_t - \phi Y_{t-1}$. (ii) Identify the process $\{U_t\}$ with justification.

   (c) Find the variance of a stationary AR(2) process.

[2+ (5+3)+5 =15]

*** xXx ***

INDIAN STATISTICAL INSTITUTE
Back Paper Examination : 2016-17
M. TECH. (QR & OR) II YEAR
Applied Stochastic Processes and Time Series Modelling

Date: 09 January 2017 | Full Marks: 100 | Duration: 3 hours

Answer all questions. Notation have usual meaning.

1. Let $N(t)$ be the number of failures of a computer system in the interval $[0, t]$. Suppose that $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda = 1$ per week. Calculate the probability that

   (a) the system operates without failure during two consecutive week.

   (b) the system will have exactly two failures during a given week, knowing that it operated without failure the previous two weeks.

   (c) less than two weeks elapse before the third failure occurs.

   [2+4+4 = 10]

2. The failures of certain device occur according to NHPP whose intensity function $\lambda(t)$ is given by:

$$\lambda(t) = \begin{cases} 0.2, & 0 < t \leq 10 \\ 0.3, & t > 10, \end{cases}$$

   where t is the age (in years) of the device.

   (a) Calculate the probability that a five-year-old device (without failure) will have exactly two failures over the next 10 years.

   (b) Knowing that the device had exactly one failure in the course the first 5 years of the 10 years considered in (a), what is the probability that this failure took place during its sixth year of use?

   [7+8 = 15]

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

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

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

   [5+8 = 13]

[P.T.O.]

4. (a) State and prove the Kolmogorov forward differential equation for CTMC. (Explain the notation used.)

   (b) Write down the generator of birth-death process.

[(3+5)+4 = 12]

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

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

[10]

6. (a) Define an weakly stationary time series.

   (b) Let $\{N(t), t \geq 0\}$ be a Poisson process with rate λ. Define a process $\{X(t)\}$ such that

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

   Verify whether $\{X(t)\}$ is weakly stationary or not.

   (c) Find an invertible process which has the following autocorrelation: $\rho_1 = 0.25$ and $\rho_k = 0$, for $k \geq 2$.

[2+8+5 =15]

7. (a) Suppose $\{X_t\}$ and $\{Y_t\}$ are uncorrelated stationary processes. Show that $\{X_t + Y_t\}$ is stationary with autocovariance function equal to the sum of the autocovariance functions of $\{X_t\}$ and $\{Y_t\}$.

   (b) Find the variance of the time series model $X_t - \phi X_{t-1} - \phi^2 X_{t-2} = Z_t$, where $Z_t \sim \mathrm{WN}(0,1)$.

[4+4 = 8]

8. (a) Describe an ARIMA (p, d, q) model.

   (b) Give the expression of forecast of x_{n+h} at time n for an ARIMA(1, 1, 1) time series model.

   (c) Consider a fitted time series model based on 100 observations $x_1, \ldots, x_{100}$ with $x_{100} = 25$:

$$x_t = 16 + 0.6x_{t-1} + Z_t + 0.8Z_{t-1},$$

   where $\{Z_t\}$ is white noise. Calculate the forecast value of x_{102} at time 100. Given that $\hat{x}_{99}(1) = 22$.

   (d) Describe three measures of forecast accuracy.

[2+3+6+6=17]

———*** xXx ***———

INDIAN STATISTICAL INSTITUTE

Semester Examination: 2016-17

(First Semester)

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

Subject Name : Reliability II

Date: 07/12/2016 Maximum Marks: 100 Duration: 3 hours

Back Paper

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

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

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

   [2x4+ 4x3=20]

2. A system consists of six subsystems that must function if the system has to function properly. The system reliability goal is 0.990. All the six subsystems have identical reliability improvement effort functions. The estimated subsystem reliabilities at the present time are 0.70, 0.75, 0.80, 0.85, 0.90, 0.95. Determine what reliability goal should be apportioned to the subsystems so as to minimize the total effort spent on the system improvement?
   If we apply the ARINC apportionment technique with the system mission time as 20 hours and the present reliability of the subsystems calculated for 50 hours, determine the apportioned reliabilities of the subsystems in order to achieve the system reliability goal of 0.990.

   [15+15=30]

3. Suppose the stress acting on a component is uniformly distributed over an interval $[S_{min}, S_{max}]$. The strength of the component has a two-parameter

Weibull distribution with parameters S_0 (shape parameter) and θ (scale parameter). Derive an expression for the reliability of the component. Let

S_{min}=10, S_{max}=30
S_0 =20, θ=30.

Find out the reliability of the component.

[15+5=20]

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

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

   [7+4+4 = 15]

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

   [(5+2) +8=15]

INDIAN STATISTICAL INSTITUTE

Back Paper Examination : 2016-17

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

Subject Name : Industrial Experimentation

Date: 11/01/2017 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. Define the terms - factor, interaction, strategy for experimentation, and nested factor.

[2 × 4 = 8]

2. Consider a $p \times p$ Latin square with rows (α_i), columns (β_k), and treatments (τ_j) as fixed factors. Write down the model. Obtain the normal equations and obtain least squares estimators of the treatment effects. How do you propose to estimate a missing observation (no derivation required)? What will be the error degrees of freedom for an approximate analysis of variance with this estimated value?

[3+12+2+2= 19]

3. (a) Consider a factorial design in two fixed factors A and B, having a and b levels respectively. Discuss the Tukey's test for detecting the presence of interaction between A and B when only one observation is taken for each treatment combination.

(b) Write down the sign table for calculating the effects in a 2^3 design with factors A, B and C. How do you compute the effect of the factor C and its sum of squares when there are n replications? Considering a 3^2 full factorial design, discuss the concepts of AB and AB^2 components of A×B interaction.

[7+(4+4+6)= 21]

4. What is partial confounding? Illustrate the concept with a suitable example. What is meant by relative information for a confounded effect? Write down the ANOVA table for your example showing only 'Source of variation' and the corresponding 'Degrees of freedom' columns.

[3+7+1+5 = 16]

5. Describe the procedure for constructing a 3^{5-2} fractional factorial design. How many treatment combinations are there in this fraction? How many generalised interactions are there? How many aliases are there for any effect?

[10+1+3+3 = 17]

2 page → 6. (a) What is a three-stage nested design?

(b) A rocket propellant manufacture is studying the burning rate of propellant from three production processes. Three batches of propellant are randomly selected from the output of each process and three determinations of burning rate are made on each batch. The result follow. Analyse the data and draw conclusions. Compute the estimates of significant model parameters including the variance component(s), if any.

	Process 1			Process 2			Process 3		
Batch	1	2	3	1	2	3	1	2	3
	19	15	15	19	23	18	14	35	25
	28	17	16	17	24	21	15	21	29
	20	14	13	14	21	17	20	24	38

[3+(10+6) = 19]

F distribution (5%) Table

$F_{0.05, v1, v2}$

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

INDIAN STATISTICAL INSTITUTE
Back Paper Examination (Semester): 2016-17
Course Name: M.Tech (QR & OR), 2nd YEAR (E & S Streams)

Subject Name: Six Sigma

Date: 12/01/2017 Maximum Marks: 100

Duration: 2 hrs.

Answer All Questions in the Answer Sheet. Work out Where It is required for Your Answer. Each Question Carries 2 Marks.

1. **Which of the following is correct about SIPOC Diagram:**
   a) A SIPOC chart gives us a high-level understanding of the process
   b) A SIPOC chart is used for identifying team members.
   c) A SIPOC chart helps in defining project benefits
   d) All of the above

2. **For a Sales Order Processing process, the enquiry of a dealer regarding the status of an order that has already been placed should be treated as**
   a) Input b) Output
   c) CTQ d) Process Measure

3. **In FMEA, the detection score should be high if**
   a) The cause happens frequently
   b) Potential failure mode is occurring frequently
   c) There is no control for that cause
   d) Process failure has severe effect on customer business

4. **Which of the following statements are TRUE about the six-sigma philosophy?**
   1. Sigma is a statistical term, which measures how far a process varies from perfection
   2. Quantitatively, Six Sigma is a stretched goal of achieving 3.4 defects per million opportunities
   3. Six Sigma is a term used to describe any cost saving achieved by the management
   4. Culturally, Six Sigma is about learning to build processes that deliver defect free output continuously

   a) Statement 1 only
   b) Statements 2 and 3 only
   c) Statements 1, 2 and 4
   d) All of the above

5. **Variation in a process exists due to**
   a) Chance causes
   b) Assignable causes
   c) Chance and assignable causes
   d) None of the above

6. **In Individual control chart, the control limits represent**
   a) The variation allowed by the customer
   b) Standard deviation of all the data
   c) Limits based on inherent process variation
   d) Allowed variation by the business head

7. **Which of the following is not a part of the project charter?**
   1. Measurable target
   2. Proposed end date
   3. Proposed solution
   4. Constraints on other CTQs
   5. Business Case
   6. Problem Statement
   7. Goal Statement
   8. Authorizing Members
   9. Solution Implementation

   a) 1 and 3
   b) 2 and 3
   c) None of the above
   d) 3 and 9

8. **There are 9 different defects that can occur on a completed time card. The payroll department collects 328 cards and finds a total of 87 defects. The Sigma Level of the process is;**

   a) 1.19 b) 3.39 c) 1.89 d) 2.13

9. **A stable, normally distributed process with specification 3.50 ± 0.05 has average of 3.50 and standard deviation of 0.025. What percent of the production violates specification?**
   a) 4.55%
   b) 5.55%
   c) 5.46%
   d) 5.00%

10. **Which of the following is NOT an expected output of the "Define" phase?**
   1. Clear objectives for the project
   2. Scope and deliverables for the project
   3. CTQ-KPIV Correlation
   4. Cause & Effect diagram
   5. Roles of team members
   6. Data Collection Plan

   a) 1 , 2 and 3
   b) 2, 4, and 6
   c) 3 ,4 and 6
   d) 5 and 6

11. **Approximately what percent of the data values are smaller than the mean, when the distribution follows Normal?**
    a) 25%
    b) 50%
    c) 75%
    d) It varies from 60% and 99%

12. **In Solution Prioritization Matrix, we first fix**
    a) Weightage for the criteria
    b) Weightage for the solutions
    c) Rank of the solutions
    d) Rank of the weightages

13. **After finding and selecting the solution, we should try to evaluate**
    a) Risks
    b) Cost benefit
    c) Implementation plan
    d) All of the above

14. **In test of hypothesis, if p-value is more than α (generally 0.05) then:**
    a) H_0 cannot be rejected
    b) H_0 is false
    c) We can't decide
    d) H_1 is false

15. **The standard deviation is calculated based on**
    a) Specification limits
    b) Target Value
    c) Observed data
    d) None of the above

16. **At Six Sigma performance level, one expects:**
    a) Max 2.4 DPMO
    b) Min 3.4 DPMO
    c) Max 4.3 DPMO
    d) Max 3.4 DPMO

17. **In Y=f(X), the negative correlation between X & Y implies:**
    a) By increasing the value of X, the value of Y will be increased
    b) By increasing the value of X, the value of Y will be decreased
    c) By increasing the value of Y, the value of X will be increased
    d) By increasing the value of Y, the value of X will almost remain unaltered

18. **Sigma rating based on continuous data depends on**
    a) Observed variation & the target
    b) SD and average
    c) Observed data, Specification
    d) All of the above

19. **In a simple linear regression set up, we find that X and Y have strong positive correlation. But the same is technologically not feasible. Also there is no computational error. Hence we conclude,**
    a) Regression analysis is not appropriate in this case
    b) Anything is possible in statistics
    c) Data collection procedure is not correct
    d) It's a breakthrough

20. **A point is out of control means –**
    a) Rejections are happening, start control
    b) Chance causes are not present
    c) Variations due to special causes are present
    d) Control limits need to be recalculated.

21. **Which of the following is FALSE?**
    1) Mean can be computed for qualitative data too.
    2) Extreme values do not affect the mode as strongly as the mean.
    3) There can be more than one mode in a distribution
    4) Extreme values affect the median as well as the mean.

    a) Only 2 and 3
    b) Only 1 and 2
    c) Only 1 and 4
    d) None of the above

22. **In order to understand the pattern in a set of continuous data for which exact timings are unknown, one will use**
    a) Histogram
    b) Run Chart
    c) Individual Control Chart
    d) None of the above

23. **In a 2-sample t-test, we deal with,**
    a) One X having two levels.
    b) Two Xs, each with two discrete levels.
    c) More than two groups
    d) None of the Above

24. **In paired t-test**
    a) We compare improvement before vs. after situations
    b) Same samples are measured before and after
    c) Observations are uncorrelated
    d) None of the above

25. **An engineer wants to study the relationship between different visual defects and various product types. The technique that will be useful for this purpose is**

a) t-test
b) F-test
c) Chi-Square test
d) ANOVA

26. **Which of the following are FALSE about the C&E diagram**

1. It unearths all possible causes for the problem at hand since it captures the views of all members
2. Creates a confusion around the problem and builds argument instead of reaching towards solutions
3. Focuses the team on the CTQ
4. Creates a consensus around the problem and builds support for obtaining solutions

a) 1 and 2
b) 2 and 3
c) 1 and 4
d) Only 2

27. **Which of the following can be classified as "Voice of Customer"?**

a) "We are happy with our work"
b) "Our response time of work orders are the best"
c) "We have competent engineers"
d) "We are the best service organization in India"
e) None of the above

28. **Which of the following is not a part of the sequence during the "data collection planning stage"?**

a) Checking for new data requirement,
b) Scrutiny of data,
c) Create data collection format,
d) Determine the sample size, and who is going to collect data

29. **A continuous variable is one which;**

a) Is counted all the time
b) Is critical to quality
c) Can be measured as a real number
d) Changes depending on the person who observes it

30. **Which of the following tools should be used to obtain VOC data?**

a) Surveys, Interviews
b) Customer Complaints
c) Service Call Rates
d) All of the above

31. **Which of the following is true about the box plot?**

1. Box plots are based on the concept of fractiles (percentiles)
2. Asterisks in the box plot indicate outliers
3. Box plots can indicate skewness in the data set
4. The whiskers in the box plots are calculated based on the first and third quartiles

a) Only 1 and 4
b) Only 1 and 2
c) Only 2 and 3
d) Only 1, 2 and 3

32. Which of the following is true for Run charts?

1. Clustering, trending and oscillations are special causes which can be identified using run charts
2. The central line in the run chart is the process mean
3. One of the tests associated with run charts is to find the special causes based on the number of runs up and down

a) Only 1 and 2
b) Only 2 and 3
c) Only 1 and 3
d) All of the above

33. A team wants a technique for obtaining a large number of possible reasons for excess variation in a dimension. They should use

a) Written and diagrammed work instructions
b) Flow charts and process maps
c) Cause and effect diagrams
d) Pareto chart

34. In SIX SIGMA projects, the initial value of SIGMA for a CTQ was 2.4 and we improved it to 3.67. All the X's identified during analysis stage are also improved. Hence, we conclude

a) Excellent work and will generate bottom-line benefit
b) Green belts and the team are not efficient
c) Improvement will not sustain in the long run
d) None of the above

35. In Six Sigma methodology we have Ys and Xs. In control phase, we establish control mechanism for

a) Ys
b) Xs
c) Both Ys and Xs
d) None of the above

36. In a one way ANOVA, which of the following is FALSE ?

1. It is a test to compare the variances
2. It is a test to compare the means of two or more groups
3. We conclude that there is significant difference between the means since the p-value of the F-test is found to be less than 0.005

a) Only 2
b) Only 1
c) 1&2
d) None of the above

37. **What is TRUE about the following ANOVA, conducted to understand the Cycle Time (CT) variation for different agents?**
One-way ANOVA: Cycle time Vs agents

Analysis of Variance for Cycle Time

Source	DF	SS	MS	F	P
Agents	2	1.20991	0.60495	157.22	0.000
Error	127	0.48867	0.00385		
Total	129	1.69857			

1) 'p'value of 0.000 indicates that there is significant difference in the mean CT of different agents
2) ANOVA was conducted to understand the difference of average cycle time of the three agents
3) 'p' value of 0.000 indicate cycle time obtained from all agents are same.

a) All of the above b) only 1 and 2
c) only 2 and 3 d) only 1 and 3

38. **Match the following:**

A-validation of the root causes	1. Define phase
B-SIPOC	2. Measure phase
C-Ranking the solutions	3. Analyze phase
D-Process capability evaluation	4. Improve phase

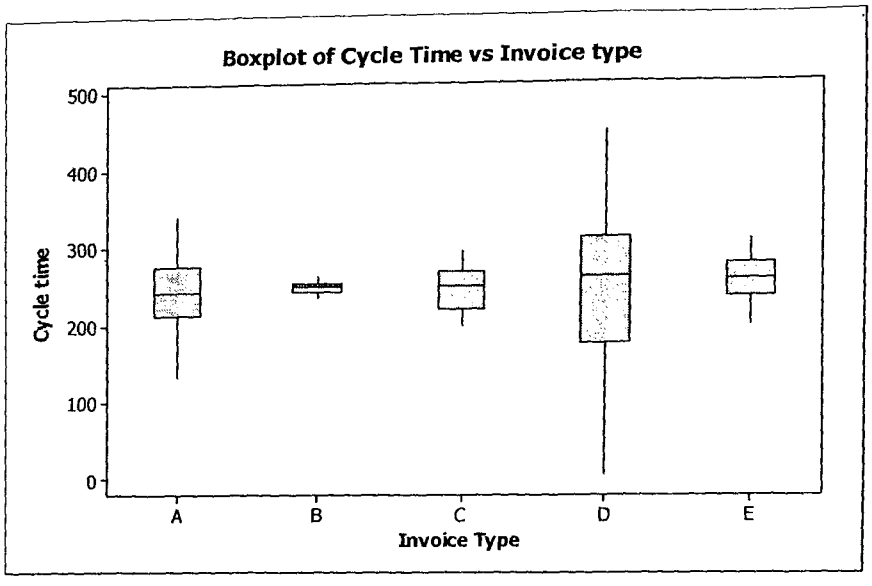
a) A3-B1-C2-D4 b) A3-B1-C4-D2 c) A4-B2-C3-D1 d) A2-B1-C3-D4

39. **For the following characteristics identify the data type- Continuous (C) /Discrete (D)**
1) Coil diameter
2) Number of imperfections per unit area

a) C – D b) C – C c) D – C d) D – D

40. **The tools used for validation of a likely cause is;**
a) Cause and Effect Diagram
b) Pareto Analysis
c) Run Chart
d) GEMBA Investigation

41. **From the Box plot it can be inferred that**



a) Invoice type differs with respect to variance of cycle time
b) Invoice type differs with respect to average of cycle time
c) both a) and b)
d) none of the above

42. A control chart is used to

a) Determine if defective parts are being produced
b) Measure process capability
c) Determine sources of process variation
d) Detect occurrence of special causes

43. The weight of the bags of a particular Raw Material is normally distributed with a mean of 200 kg. We know that 95.46% of the bags weigh between 190 kg and 210 kg. Which of the following is likely to be the standard deviation of the distribution?

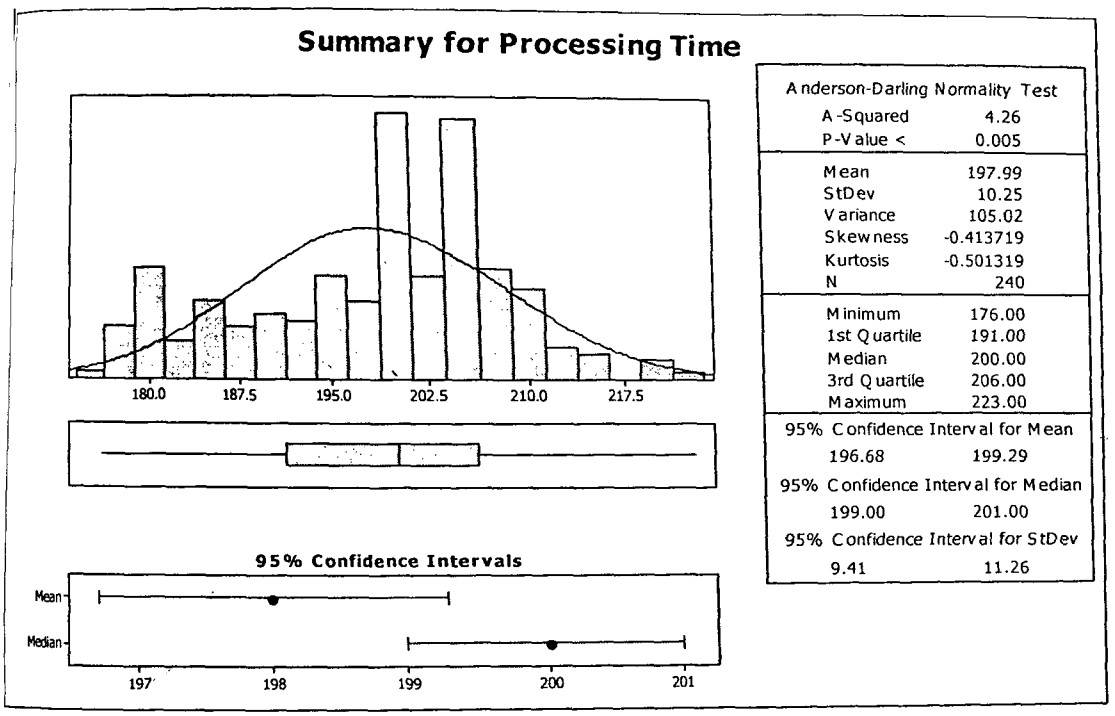
a) 8kg b) 2kg c) 4kg d) 5kg

44. Which three of the following four techniques could easily be used to display the same data?

i. Dotplots
ii. Boxplots
iii. Scatter Diagrams
iv. Histograms

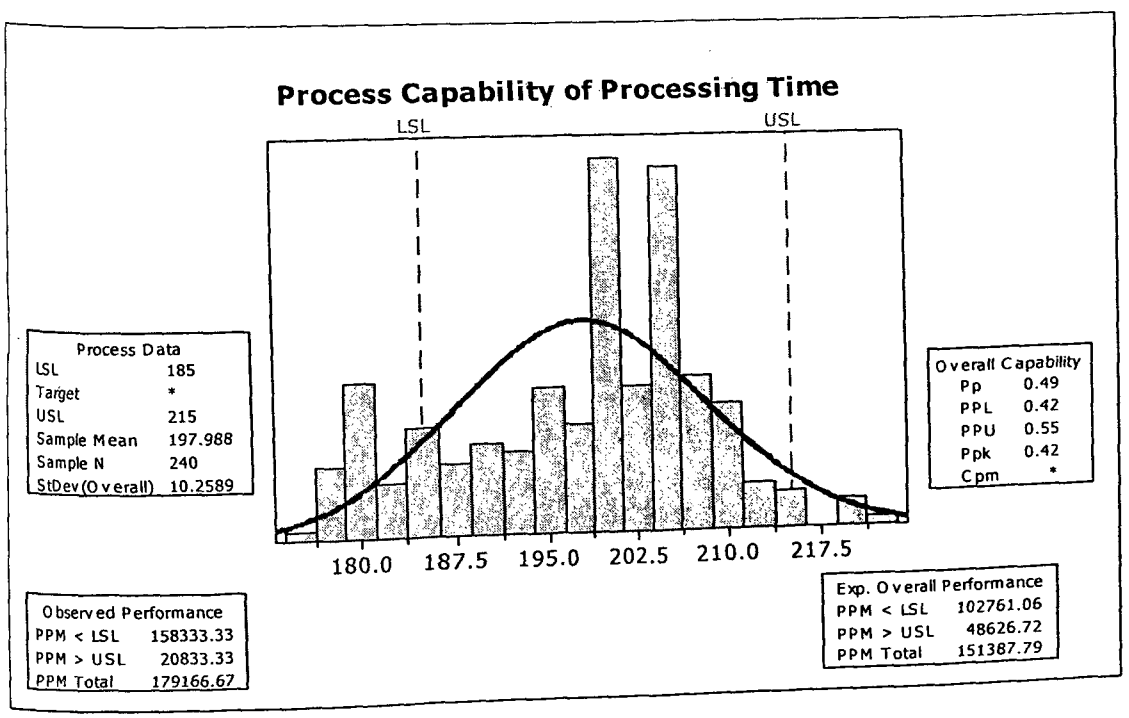
a) i, ii and iii only
b) i, ii and iv only
c) i , iii and iv only
d) ii, iii and iv only

45. Summary statistics of Processing Time are given in the following diagram. We can infer that



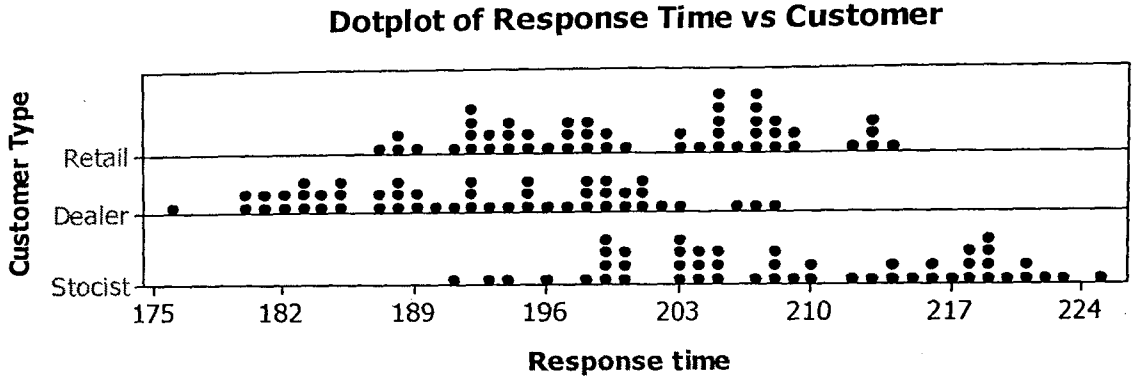
a) The process follows Normal distribution
b) The process does not follow Normal distribution
c) Not possible to comment with the above information
d) None of the above

46. The approximate Sigma Level of the process is



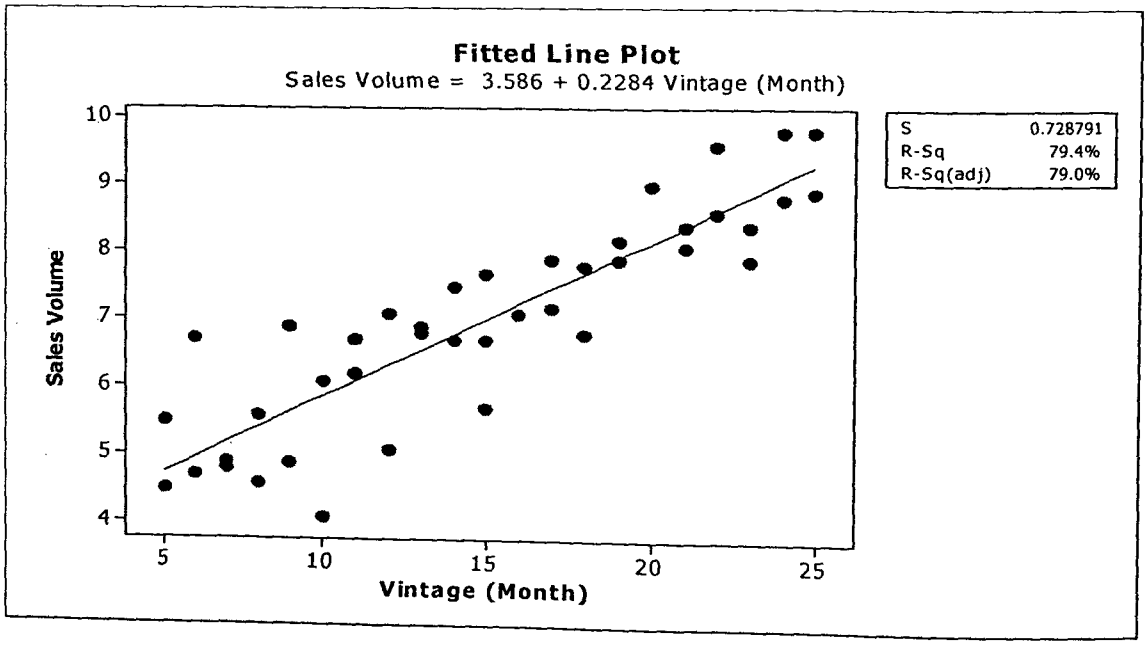
a) 2.8 b) 2.4 c) 1.5 d) 1.4

47. From the following Dot Plots one can infer that



a) Response Time differs with respect to variance among Customer types
b) Response Time differs with respect to average among Customer types
c) Response Time does not differ on any aspect among Customer types
d) None of the above

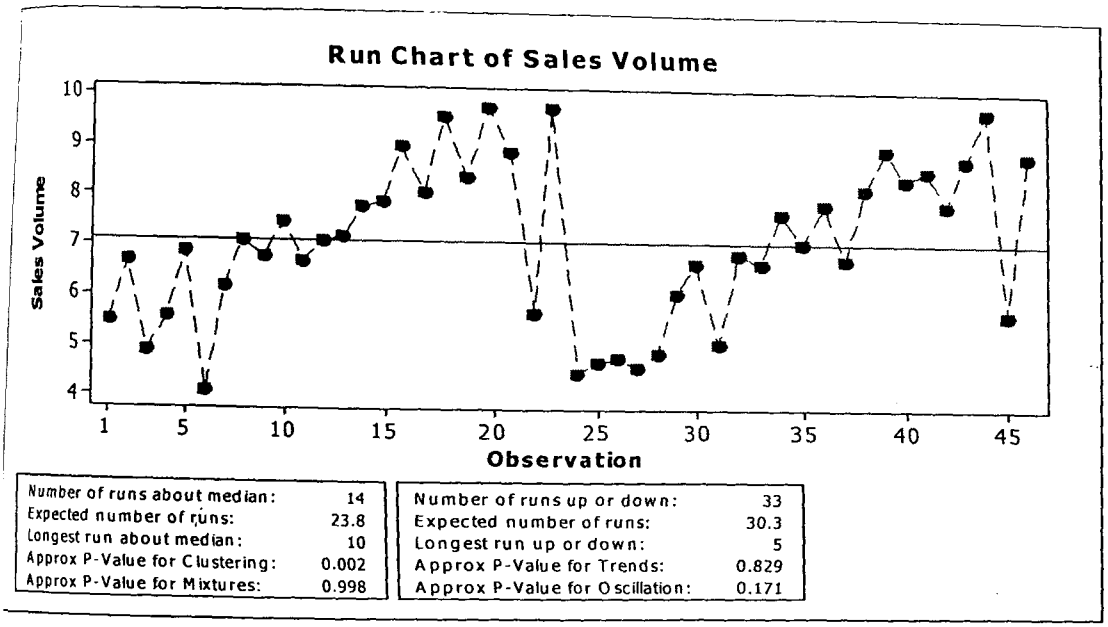
48. An organization wants to find out whether vintage of employees affects the sales volume or not. Data are collected and plotted in a scatter diagram



From the above analysis we can conclude that

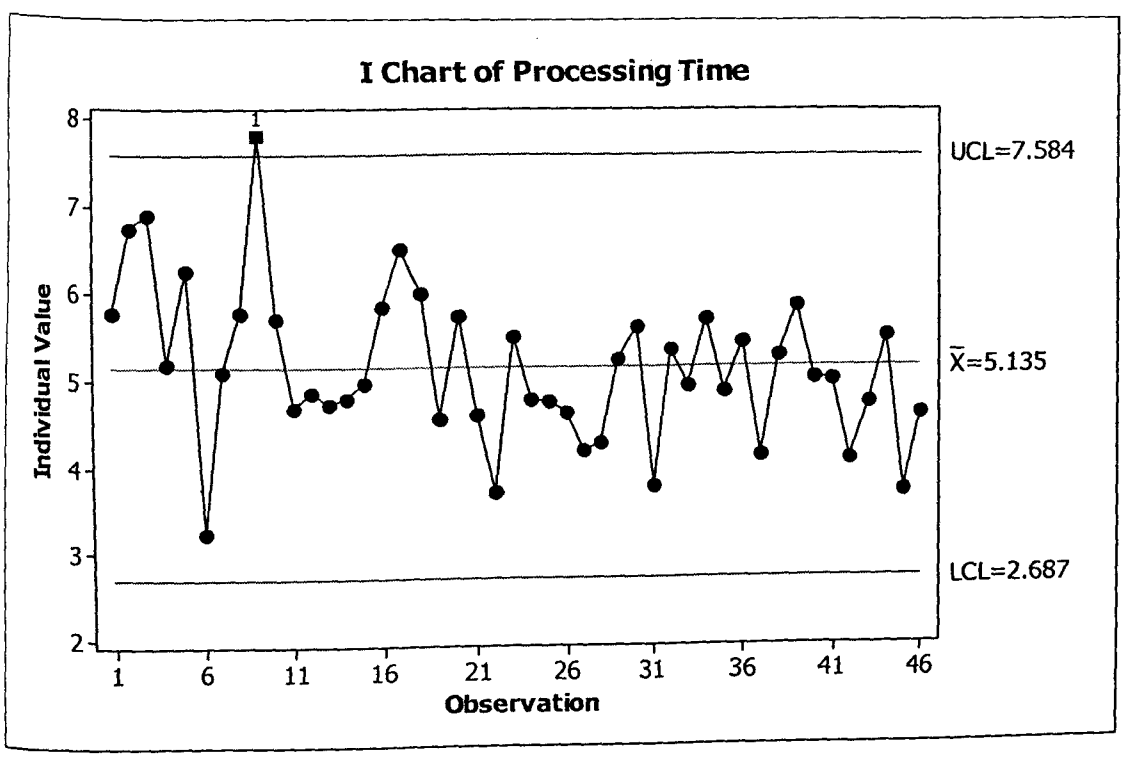
a) Vintage affects sales volume
b) Vintage affects sales volume Significantly
c) Vintage does not affect sales volume
d) Results are inconclusive.

49. Sales volume of a Dealer is collected over a few weeks and is shown in the following Run Chart. From the chart one can conclude that



a) The variation is due to assignable causes only
b) The variation is due to chance causes only
c) The variation is due to both chance & assignable causes
d) The variation is due to either chance or assignable cause

50. Data collected on processing time are analyzed through Individual chart. From the chart one can conclude that



a) The process is unstable
b) The MR chart would be uninformative
c) The process is having a trend
d) The measurement/sampling is an issue.