

- N. B. :** (1) All questions are **compulsory**.
 (2) Use of simple calculator is **allowed**.
 (3) **Figures** to the **right** indicate **full** marks.
 (4) Answers to the **two** sections should be written on **separate** answer books.

Section I

1. (a) (i) For a sample of 10 measurements it was found that :- 4

$$\sum_{i=1}^{10} x_i = 30 \text{ and } \sum_{i=1}^{10} x_i^2 = 266$$

Find the standard deviation of the distribution.

- (ii) Find the median and the lower and upper quartiles for the following data :
 8, 7, 1, 4, 6, 6, 4, 5, 7, 6, 3, 0

- (b) The probability that a family owns a house is 0.76. The probability that a family owns a house and is a married couple is 0.69. Find the conditional probability that a randomly selected family is a married couple given that it owns a house. 2
 (c) Two methods **A** and **B** are available for teaching a certain industrial skill. The failure rate is 20% for **A** and 10% for **B**. However **B** is more expensive and hence is used only 30% of the time. (**A** is used other 70%) A worker was taught the skill by one of the methods but failed to learn it correctly. What is the probability that she was taught by method **A** ? 4

2. (a) A claim size distribution is modeled using a simple distribution with density 4

$$\text{of the form } f(x) = \begin{cases} k(50-x) & 0 \leq x \leq 50 \\ 0 & 0 < x < \infty \end{cases}$$

- (i) Find k
 (ii) $p(x > 25)$
 (iii) Probability that an individual claim size is less than 30 given that it is greater than 25.
 (b) Assume that 40% of the policy holder in a certain metropolitan area have type **A** blood. If the distribution of the blood donors among the policy holders entering a blood bank on any given day is considered random : 3
 (i) Find the distribution of x , the number of donors entering a blood bank on a given day, until the first type **A** donor is encountered.
 (ii) Also find the mean and standard deviation of x .
 (c) Let y denote a Poisson random variable with mean λ . Find the probability generating function for y , and use it to find $E(y)$ and $v(y)$. 3

3. (a) The joint p.d.f. of (x, y) is given below 3

$$f(x, y) = \begin{cases} 4xy & 0 < x < 1 \\ & 0 < y < 1 \\ 0 & 0 < x < \infty \end{cases}$$

Test whether x and y are independent.

- (b) Suppose that the amount of time a customer spends at a cash counter in a certain office has an exponential distribution with a mean of six minutes. Find : 3
 (i) The probability that a randomly selected customer will spend more than 12 minutes.
 (ii) The conditional probability that the customer will spend more than 12 minutes in the cash counter given that the customer has been there for more than six minutes.

[TURN OVER

- (c) Let $S = x_1 + x_2 + \dots + x_N$, where N represents the random number of claims received during the time period, x_i represents the size of the i^{th} claim and S represents the aggregate claim. Assuming x_i 's are iid and are also independent of N . Show that –

$$(i) E(s) = \mu_N \mu_x$$

$$(ii) \text{Var}(s) = \mu_N \sigma_x^2 + \sigma_N^2 \mu_x^2.$$

Section II

4. (a) Let x_1, \dots, x_n denote a random sample from $U(0, \theta)$ with p.d.f.

$$f(x, \theta) = \frac{1}{\theta}, \quad 0 < x < \theta, \theta > 0$$

$$= 0, \quad \text{otherwise}$$

- (i) Obtain MLE of θ .
 (ii) Obtain moment estimator of θ .

Is it unbiased? Find its variance.

- (b) Let x_1, \dots, x_n be a random sample from $p(\lambda)$ distribution. Obtain MLE of λ and CRLB of variance of any unbiased estimator of λ . Obtain moment estimator of λ . Is it unbiased? Find its variance.

5. (a) A certain stimulus administered to each of 12 patients resulted in the following increases of blood pressures.

5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4

can it be concluded that the stimulus will be in general, accompanied by an increase in blood pressure?

- (b) The following table shows the number of persons in a random sample of 210 listed according to the day of the week on which they prefer to do their grocery shopping :-

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
No. of Persons	9	17	12	26	36	69	41

Using $\alpha = 0.025$ test the null hypothesis that the proportion of persons who prefer to do their grocery shopping on a particular day is the same for all days of the week.

6. (a) A company claims that its medicine Brand **A** provides faster relief than another company's medicine Brand **B**. A researcher tested both brands of medicine on two groups of randomly selected patients. The results of the test are as follows :

Brand	Sample size	Mean of relief times (in minutes)	Std. deviation of relief times (in minutes)
A	25	44	11
B	22	49	9

- (i) Construct a 99% confidence interval for the difference between the mean relief times for the two brands of medicine (in minutes).
 (ii) Test at the 1% level of significance whether the mean relief time for brand **A** is less than that for brand **B**. State the assumptions.
 (iii) Test at 5% level of significance whether the two population variances are same.
- (b) In a survey 66% of the respondents agreed that police should be allowed to collect DNA information from suspected criminals. In a recently taken random sample of 500 Americans 305 held this view.
- (i) Using a 5% level of significance, can you conclude that the current proportion who hold this view differs from 66%.
 (ii) Construct 95% confidence interval for the population proportion p who hold this view.

(3 Hours)

[Total Marks : 60

- N.B.** (1) All questions are **compulsory**.
 (2) Use of **simple** calculators is **allowed**.
 (3) **Figures** to the **right** indicate **marks**.
 (4) Answer to the **two** sections should be written on **separate** answer books.

Section I

1. Explain :— 10
 - (a) Single Exposure Test including examples.
 - (b) Copy Testing — Dummy magazine test.
2. Design "Segmentation Survey Questionnaire" to gather attitudes and behaviour to look for customer segments. 10
3. Explain the entire process along with its importance of report preparation and presentation. 10
4. Design questions for Advertising Research. 10

Section II

5. A company has been supplying power supply generator to companies manufacturing computers, washing machines, TVs. It has used cluster analysis to find out segments based on needs of its customers. It has used 7 point scale where 7 means very important and 1 means not very important. The needs used are importance of manufacturer's assistance in developing product, importance of product warranty, importance of newest technology, importance of price, importance of well known manufacturer and reliable distribution. Following is the output. Interpret the output :— 10

Iteration History (a)

Iteration	Change in Cluster Centers	
	1	2
1	1.832	3.212
2	.261	.755
3	.000	.000

Final Cluster Centers

	Cluster	
	1	2
Importance of Mfr. Assistance	6	4
Importance of Product Warranty	5	3
Importance of Newest Technology	3	5
Importance of Price	3	5
Importance of Well-known Mfr.	3	4
Importance of Reliable Distribution	4	5

[TURN OVER

ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Importance of Mfr. Assistance	227.010	1	.214	228	1060.171	.000
Importance of Product Warranty	180.053	1	.516	228	348.880	.000
Importance of Newest Technology	127.135	1	.686	228	185.399	.000
Importance of Price	108.539	1	1.000	228	108.529	.000
Importance of Well-known Mfr.	46.575	1	.688	228	67.687	.000
Importance of Realiable Distribution	40.488	1	1.055	228	38.374	.000

Number of Cases in each Cluster

Cluster 1	184.000
2	46.000
Valid	230.000
Missing	.000

6. (a) Listed are 6 questions taken from a survey of dealers selling heavy equipments manufactured by a manufacturer. Marketing Researcher used factor analysis on 6 statements and extracts two factors. Varimax rotated component matrix is given below :— 10

	Factor 1	Factor 2
Q. 1 Overall we are satisfied with this manufacturer	0.703	0.286
Q. 2 We feel good about working relationship we have with this manufacturer	0.642	0.309
Q. 3 We look forward to continuing our business relationship with this manufacturer	0.743	0.251
Q. 4 This manufacturer is always ahead of his competitors	0.243	0.714
Q. 5 This manufacturer is a market leader.	0.356	0.639
Q. 6 This manufacturer is achieving its sales potential	0.332	0.624

What inference can he draw from above table ?

- (b) Differentiate between Orthogonal and Oblique Factor Rotation methods in factor analysis. Explain the Varimax method of factor rotation. 10

Regression and Linear model

BB-6529

(3 Hours)

Jan. 2012

Total Marks : 60

- N.B. :** (1) All questions are **compulsory**.
 (2) Use of **simple** calculator is **allowed**.
 (3) **Figures** to the **right** indicate **full** marks.
 (4) Answer to the **two** sections should be written on **separate** answer-books.

Section I

1. Assume regression model $Y = \beta_0 + \beta_1 X + e$ and n observations.
 - (a) Prove that Regression S.S. = $\hat{\beta}_1^2 \sum (x_i - \bar{x})^2$. 3
 - (b) Data on 10 observations gives :- 7
 $\sum (Y_i - \hat{Y}_i)^2 = 3.267$, $\sum Y_i = 63.3$, $\sum Y_i^2 = 423.49$, $\sum x_i = 139$, $\sum x_i^2 = 2239$.
 Estimate the parameters β_0 and β_1 . Find :-
 (i) Var ($\hat{\beta}_0$), (ii) S.e. ($\hat{\beta}_1$) (iii) Coefficient of determination.
2. (a) If there is one response variable, p predictor variables and n observations, show the data set in a tabular form. 2
 (b) The average monthly returns of BSE sensx during 10 months along with those of stock A and stock B are available. Formulate a model to describe average monthly returns of BSE stock in terms of that of stock A and stock B. Assume linear relation. 3
 (c) State the assumptions that are required in regression analysis. Briefly explain how to check violation of assumptions. 5
3. (a) How principal components analysis helps to detect multicollinearity? Eigen values of correlation matrix of three predictor variables are, $\lambda_1 = 2.952$, $\lambda_2 = 0.040$, $\lambda_3 = 0.008$. Use two criteria based on eigen values to check the presence of multicollinearity. 2
 (b) How do you model 0 or 1 dependent variable using logistic regression? The 95% confidence interval for odds ratio corresponding to three predictor variables are given below. Comment about effect of each variable on odds ratio :-
 $X_1 : (0.77, 2.51)$ $X_2 : (0.97, 1.48)$ $X_3 : (0.01, .93)$. 5
 (c) Why is transformation of variables necessary? State the transformations for the following models :- 3

$$(i) Y = \alpha x^\beta \quad (ii) Y = \alpha e^{\beta x} \quad (iii) Y = \frac{X}{\alpha x - \beta}$$

[TURN OVER

Section II

4. (a) Write down the one-way classification model with all its assumption. Further state all the SAS commands for this model. 10
- (b) Fifteen fourth grade students were randomly assigned to three groups in order to experiment with three different methods of teaching arithmetic. At the end of the semester, the same test was given to all fifteen students. Obtain 95% confidence intervals for the three methods ($\alpha_1 - \alpha_2$) and ($\alpha_1 - \alpha_3$) for the following data :-

$$\sum_{i=1}^3 \sum_{j=1}^{n_i} y_{ij}^2 = 80709, \quad \sum_{i=1}^3 \frac{y_{i.}^2}{n_i} = 78336.2$$

$$y_{..} = 1081, \quad n_1 = n_2 = n_3 = 5, \quad F_{1,12,.05} = 4.75$$

Where $y_{ij} = \mu + \alpha_i + e_{ij}$, $i = 1, 2, 3$, α_i deriotes $j = 1, 2, -n_i$, effect of the method.

5. (a) Explain Nested Models with examples. Specify its ANOVA. 10
- (b) Two laboratories were used to determine the purity of a chemical compound. Within each of the laboratories three technicians were used to carry out the analysis.

Let p = No. of laboratories = 2

q_i = No. of technicians in each laboratory, $i = 1, 2$

$q_1 = q_2 = 3$

$$\sum_{i=1}^2 \sum_{j=1}^{q_i} \sum_{h=1}^{r_{ij}} y_{ijh}^2 = 19529, \quad y_{...} = 979$$

$$\sum_{i=1}^2 \frac{y_{i..}^2}{R_{ij.}} = 17817.81, \quad \text{where } R_{ij.} = \text{Total no. y observation in } i\text{th lab.} = 27$$

$$\sum_{i=1}^p \sum_{j=1}^{q_i} \frac{y_{ij.}^2}{r_{ij.}} = 17857.44, \quad Q = q_1 + q_2 = 6, \quad R_{..} = 54$$

$$F_{1,48,.05} = 4.05 \quad F_{4,48,.05} = 2.58$$

Test the hypothesis that there is no difference in laboratories and technicians within the laboratories at 5% level of significance.

6. (a) Write down the three way model with all its assumption. Further write down its ANOVA table. 10
- (b) Write a short note on Q - Q plot and Tukey's test.

- N. B. :**
- (1) All questions are **compulsory**.
 - (2) Use of simple calculators is **allowed**.
 - (3) **Figures** to the **right** indicate **marks**.
 - (4) Answer to the **three** sections should be written on **separate** answer books.

Section I

1. (a) Explain the circumstances under which it is advisable to employ :— 10
 - (i) Qualitative Methods,
 - (ii) Causal Models and
 - (iii) Time Series Techniques.
- (b) ABC Telefilms Ltd. wants to find out whether Family Drama or Reality Show or Quiz Contest or an entirely new programme will become the most popular TV show format two years from now. Prepare a plan to carry out this research study.
2. Which of the following five statements are correct ? Mention your opinion as 'CORRECT' or 'WRONG' against each statement and give detailed reasons to support your opinion. 10
 - Statement Number 2.1 :** A quantitative model can not be considered appropriate unless error analysis is carried out.
 - Statement Number 2.2 :** Moving averages of period 3 (three) should effectively smoothen the quarterly sales figures of travel goods.
 - Statement Number 2.3 :** Strong correlation between two variables does not necessarily mean a cause and effect relationship between the two.
 - Statement Number 2.4 :** In time series analysis, it is safer to employ the additive model than the multiplicative model.
 - Statement Number 2.5 :** The classical decomposition model of time series is not recommended for long term forecasting.
3. (a) Describe the 'Six Thinking Hats Technique' for making a decision in complex real life situations. 10
- (b) How will you use it to decide whether a cash-rich software company should build and operate a shopping mall ?

Section II

4. Define problem of Economic Order Quantity with known demand. The shortages are allowed. Rate of replenishment of inventory is finite. Determine optimum production quantity for each production run stating the conditions and diagram. 10
5. For a fixed order quantity system find out the E. O. Q., optimum buffer stock, normal lead time consumption and reorder level with the following data ; 10

Annual consumption in units	= 4900
Cost in Rs. of one unit	= 1
Cs (set up cost) in Rs. per Production Run	= 12
C1 (holding cost) in Rs. per unit	= 0.24
Past lead times	= 10 days, 20 days, 8 days, 9 days, 25 days and 12 days.

Section III

6. The Ward Company sold 2,00,000 units of its product at Rs. 20 per unit. Variable costs are Rs. 14 per unit. Fixed costs are incurred uniformly throughout the year and amount to Rs. 7,92,000. 10
 - (a) Required to determine the following :—
 - (i) The Break even point for this product. (In terms of Rupees and Units).
 - (ii) Margin of safety for the company. (In terms of Rupees and Units).
 - (iii) The number of units that must be sold to earn a net profit of Rs. 60,000 for the year.
 - (b) The number of units that must be sold to earn a net profit of Rs. 90,000 if the selling price has been increased to Rs. 22/-.

Statistical Process Control
January 2012

BB-6547

Con.- 7053-11.

(3 Hours)

[Total Marks : 60

- N.B. :** (1) All questions are compulsory.
 (2) Figures to the right indicate marks.
 (3) Answers to the two sections must be submitted on separate answer books.
 (4) Calculators are allowed.

SECTION 1

Instructions:

- (1) Use Statistical Software & Excel along with six sigma excel template
 (2) Write down Null & Alternative hypothesis for the test with name/s of the test/s used, basis of test and conclusion along with output of session window

Q1. Explain in detail following :

- (a) SIPOC of your Project
 (c) FMEA Concept

(b) Steps of 5 S

(d) Filter Score of Team Meeting.

(10)

Q2.

- (a) A Company wants to understand which Factory is better & reliable w.r.t. quality in order to mfg. premium brand. (Higher the number – Better the quality) (10)

Factory A	175	165	1160	154	169	170	182	177	159	162	176	163
Factory B	165	180	152	174	179	180	172	188	168	160	182	173

- (b) Given below efficiency % of a large Manufacturing unit. Kindly help in setting improvement Target for their Project Team.

Month	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
Wastage %	83.4	72.1	82.9	93.1	72.4	81.9	92.2	82.8	78.6	88.0	76.5	79.0

Q3. Following 4 m/c factors determine the ceramic strength. Conduct the DOE with replication & write down the regression equation based on p-value. (10)

F1 : Temp (cold and hot)

F2 : Feed Rate (slow and high)

F3 : Operating Skills (Average and Excellent)

F4 : Shift (Day & Night)

Run	F1	F2	F3	F4	Ceramic Strength	Random Order	Run	F1	F2	F3	F4	Ceramic Strength	Random Order
1	-1	-1	-1	-1	430.45	17	17	-1	-1	-1	-1	357.34	12
2	1	-1	-1	-1	472.48	30	18	1	-1	-1	-1	370.80	1
3	-1	1	-1	-1	452.14	14	19	-1	1	-1	-1	360.55	4
4	1	1	-1	-1	416.93	8	20	1	1	-1	-1	388.04	23
5	-1	-1	1	-1	453.67	32	21	-1	-1	1	-1	335.19	2
6	1	-1	1	-1	392.14	20	22	1	-1	1	-1	336.17	28
7	-1	1	1	-1	442.98	26	23	-1	1	1	-1	351.67	11
8	1	1	1	-1	419.26	24	24	1	1	1	-1	358.31	9
9	-1	-1	-1	1	241.58	10	25	-1	-1	-1	1	192.90	25
10	1	-1	-1	1	225.52	16	26	1	-1	-1	1	284.41	21
11	-1	1	-1	1	228.76	27	27	-1	1	-1	1	167.66	6
12	1	1	-1	1	318.23	18	28	1	1	-1	1	260.84	7
13	-1	-1	1	1	194.72	3	29	-1	-1	1	1	142.11	5
14	1	-1	1	1	160.37	19	30	1	-1	1	1	93.22	13
15	-1	1	1	1	178.51	31	31	-1	1	1	1	135.52	22
16	1	1	1	1	241.47	15	32	1	1	1	1	196.73	29

[TURN OVER

Con. 7053-BB-6547-11.

2

SECTION 2

Q. 4(a) Suggest appropriate QC tool from 7 New QC tools that can be used for following problems: (5)

Sr.	Problem
1	Company having multiplexes across India would like to understand relationship between increases in ticket prices (10%, 15%, 20% and 25%) against various types of cities/towns.
2	NGO working on environmental issues visit a village where there are many environmental issues. They want villagers to be part of their environment improvement programme and collect suggestions from all.
3	A toy store was aiming to increase sales while improving the satisfaction of its customers with the toys that it sold. It would like to measure both the initial appeal (which related to actual purchase) and the longer term satisfaction (which related to company image) of a range of toys for boys aged 5 to 10, both being scored on a one-to-ten scale.
4	A production group in an electrical goods manufacturer measures and maps out the tasks required to build a toaster. They redesign and reallocate tasks to reduce the critical path time.
5	The design department and the production department of a company had continuous issues of bad changes to design by production department and designs provided by design department to be impractical to build. They felt there could be common causes to these issues.

Q. 4(b) An airlines company would like to analyze and prioritise the quality complaints received from its customers. The complaint data is as below: (5)

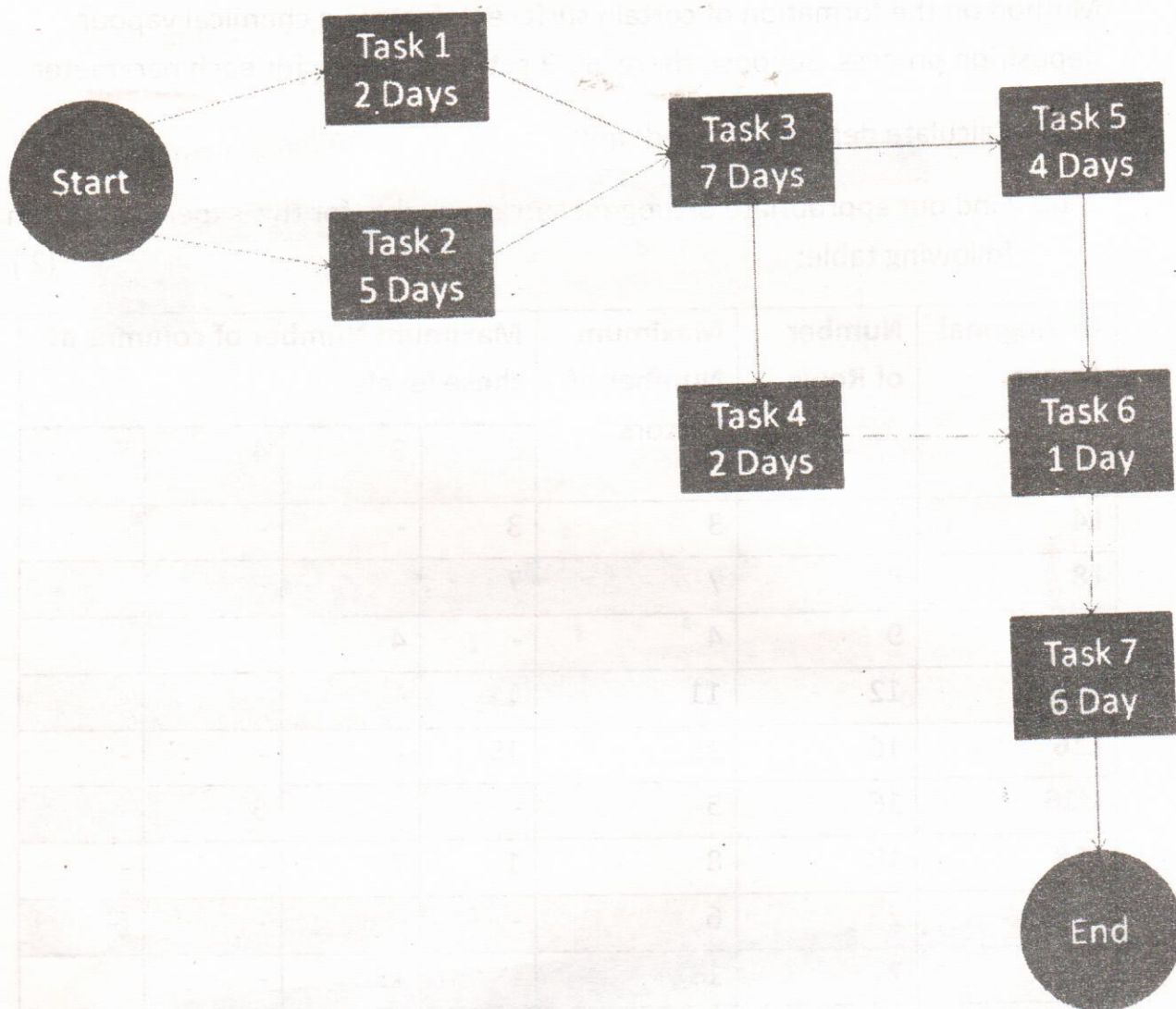
Type of complaint	Number
Baggage delay	23
Missed connections	15
Lost baggage	7
Poor cabin service	3
Ticketing error	2

Which QC tool can be used to represent this? Use appropriate QC tool, draw appropriate analysis and write your findings.

Con. 7053-BB-6547-11.

3

Q. 5. Following activity diagram represents project plan of road repairs to be done before monsoon. Answer following questions. (10)



- Calculate early start and finish time for each task.
- Calculate late start and finish time for each task.
- Identify Critical path.
- When the project must start if required to be completed on 31st May?
- When can we start Task 4 (latest) without impacting overall schedule?

[TURN OVER

Con. 7053-BB-6547-11.

Q. 6. (a) Consider a project where we are interested in determining the effect of four process parameters, Temperature, Pressure, Settling Time and Cleaning Method on the formation of certain surface defects in a chemical vapour deposition process. Suppose there are 3 settings chosen for each parameter.

- (i) Calculate degrees of freedom? (2)
- (ii) Find out appropriate orthogonal array suitable for this experiment from following table: (2)

Orthogonal Arrays	Number of Rows	Maximum Number of Factors	Maximum Number of columns at these levels			
			2	3	4	5
L4	4	3	3	-	-	-
L8	8	7	7	-	-	-
L9	9	4	-	4	-	-
L12	12	11	11	-	-	-
L16	16	15	15	-	-	-
L'16	16	5	-	-	5	-
L18	18	8	1	7	-	-
L25	25	6	-	-	-	6
L27	27	13	-	13	-	-
L32	32	31	31	-	-	-
L'32	32	10	1	-	9	-
L36	36	23	11	12	-	-
L'36	36	16	3	13	-	-
L50	50	12	1	-	-	11
L54	54	26	1	25	-	-
L64	64	63	63	-	-	-
L'64	64	21	-	-	21	-
L81	81	40	-	40	-	-

Con. 7053-BB-6547-11.

5

Q. 6(b) In a case study there are four 2-level factors A, B, C and D. We want to estimate their main effects and also the interaction A x B, B x C and B x D. Calculate degrees of Freedom. Use following L8 orthogonal array and interaction table and draw a table of experiment layout. Show columns representing interaction between A x B, B x C and B x D. (6)

L8 – Orthogonal Array

	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
	A	B	C	D	E	F	G

Interaction table for L8

	1	2	3	4	5	6	7
1	(1)	3	2	5	4	7	6
2		(2)	1	6	7	4	5
3			(3)	7	6	5	4
4				(4)	1	2	3
5					(5)	3	2
6						(6)	1
7							(7)

(3 Hours)

[Total Marks : 60

- N. B. : (1) All questions are **compulsory**.
 (2) **Figures** to the **right** indicate **marks**.
 (3) Answer to the **two** sections must be submitted on **separate** answer books.
 (4) Calculators are **allowed**.

SECTION I

Q.1. (a) Describe use of i) Descriptive Statistics and ii) Graphical Method in summarizing data having multivariate observations. (05)

(b) Let $X' = (X_1, X_2, X_3, X_4)$ be a four variate random vector with mean vector $\mu' = (3 \ 5 \ 2 \ 4)$ and variance-covariance matrix.

$$\Sigma = \begin{bmatrix} 3 & 0 & 5 & 0 \\ 0 & 1 & 0 & 3 \\ 5 & 0 & 4 & -1 \\ 0 & 3 & -1 & 1 \end{bmatrix}$$

If $A = (1, 3, -2, 1)$.

(1) Obtain mean and variance of $A X$

(2) Find correlation co-efficient between :

(i) X_1 and (X_2, X_4) (ii) X_3 and X_1 (iii) X_2 and X_3 (05)

Q.2. (a) Consider the following data

$$X = \begin{bmatrix} 3 & 2 & 2 \\ 0 & 1 & 5 \\ 4 & 2 & 6 \end{bmatrix}$$

$$\text{If } b = \begin{bmatrix} 3 \\ 3 \\ -2 \end{bmatrix} \text{ and } c = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

Obtain sample mean & sample variance of $b' X$ and $c' X$. Obtain covariance between $b' X$ and $c' X$. (04)

(b) Let X_1, X_2, X_3, X_4 and X_5 be independent and identically distributed random vectors with mean vector μ & covariance Matrix Σ . Obtain mean vectors and covariance Matrix for

(i) $\frac{1}{10} X_1 + \frac{1}{10} X_2 + \frac{1}{10} X_3 + \frac{1}{10} X_4 + \frac{1}{10} X_5$

(ii) $X_5 - X_4 + X_3 - X_2 + X_1$

(iii) $\sum_{i=2}^5 (X_i - X_{i-1})$

[TURN OVER

Q.3. (a) State p. d. f. of $X \sim N_p(\mu, \Sigma)$

$$\text{Partition } X = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix} \quad \& \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

And $|\Sigma_{22}| > 0$. State the conditional p. d. f. of $X_{(1)}$ given $X_{(2)} = x_{(2)}$ (4)

(b) State properties of Multivariate Normal Distribution. (4)

(c) Explain in brief 'Transformations to near normality'. (02)

Q.4. (a) A random sample of size 'n', X_1, X_2, \dots, X_n is selected from $N_p(\mu, \Sigma)$.

Obtain the likelihood ratio test to test the hypothesis

$H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$. State its relation to the Hotelling T^2

Statistics. State $100(1-\alpha)\%$ simultaneous confidence interval for

$\mu_i - \mu_k, 1 \leq i, k \leq n, i \neq k$. (05)

(b) Describe the multivariate multiple regression model. Indicate method of estimating parameters.

SECTION II

Q.5. (a) Write down the orthogonal factor model with m factors. (04)

(b) Explain the use of scree plot in principal component analysis and factor analysis. (03)

(c) Obtain $(1-\alpha)\%$ confidence interval for variance of the first principal component. (03)

Q.6. (a) Define the first two pairs (U_1, V_1) and (U_2, V_2) of canonical variables. What is $V(U_1)$ and $\text{cov}(U_1, V_2)$? (4)

(b) Describe the problem of classification with respect to two populations. Define the distance between two Normal populations with common dispersion matrix. (03)

(c) Explain how to test the hypothesis that distance between two populations is Zero. (03)

.....

medical stats

Jan 2012

Con.- 7054-11.

BB-6550

(3 Hours)

[Total Marks : 60

- N.B. :** (1) All questions are **compulsory**.
 (2) **Figures** to the **right** indicate marks.
 (3) Calculators are **allowed**.

1. Briefly explain the application of following Concepts / tests / methods in analysis of clinical research data. [Any 5 of 7, 2 Marks each] (10)

- ANCOVA
- Wilcoxon rank sum test
- Logistic regression
- Fisher's Exact Test
- Randomization
- Box-Whisker Plot
- Binomial Test.

2. Describe process flow of a Clinical trial and discuss the role of a statistician in a clinical trial project. (10)

3. Answer following questions (5 marks each) (10)

- Explain P-value, confidence interval, type II error and power of a test
- ICH E9 Guideline.

4. An investigator wants to compare effect of two Drugs (Drug A and B) in increasing haemoglobin value of Female patients suffering from Iron Deficiency Anaemia. (10)

All female patients to be recruited will be of similar age and are expected to have similar low baseline Haemoglobin and serum Iron levels. (Normal values for Hemoglobin (Hb) are between 12 gm/dL to 18 gm/dL and Serum Iron normal values are between 60-170 mcg/dL). Investigator does not expect any other factors other than treatment to influence haemoglobin levels in such patients.

End points of the trial are Haemoglobin value at end of 2 months of treatment. Serum iron values at the end of 4 months of treatment. Investigator wants to compare mean Serum Iron and mean Serum haemoglobin values across two groups.

He wants to prove that patients who receive treatment with Drug A have mean Hemoglobin value that is greater by 1 gm/dL as compared to those treated with drug B at the end of two months. He also wants to show that Serum Iron levels at end of 4 months will be greater in patients treated with Drug A as compared with Drug B.

This will help him to promote Drug A as treatment of choice at his institution. As a statistician for this study, please provide following information.

- a. Hypothesis for the study
 - b. Information that you will need for calculating sample size for this study
 - c. Statistical test that you intend to apply to test investigator's claim
 - d. Statistical considerations involved in analysis of this data
5. What are non-parametric tests and in which situation would you use non-parametric tests to analyse continuous data. Which non parametric tests would you use to compare (10)
- a. Change from baseline blood pressure in a group of patients
 - b. Change from baseline blood pressure in patients receiving 4 different treatments.

[TURN OVER

Con. 7054-BB-6550-11.

2

6. (Multiple Choice Question Answers –EACH QUESTION HAS ONE CORRECT CHOICE [2 marks each]) (10)

Select correct choice and write the number against the choice in your answer book for the following questions ; as shown in the example below.

(a) Sample question text _____ ?

- (i) Choice 1 (iii) Choice 3
(ii) Choice 2 (iv) Choice 4.

Answer : (a) (iv).

- i) Which one of the following output is produced when the program below is submitted?

```
Proc sort data=save.test;
```

```
By abc;
```

```
Run;
```

- a. The descriptor portion of dataset SAVE.TEST
- b. The descriptor portion of dataset WORK.TEST
- c. The data portion of dataset SAVE.TEST
- d. None of the above .

- ii) Which one of the following output is produced when the program below is submitted?

contents
Proc ~~sort~~ data=save.test;

```
By abc;
```

```
Run;
```

- e. The descriptor portion of dataset SAVE.TEST
- f. The descriptor portion of dataset WORK.TEST
- g. The data portion of dataset SAVE.TEST
- h. None of the above .

- iii) Identify the missing:

```
Data new.exam;
```

```
_____ exam;
```

```
Where var1 > var2;
```

```
Run;
```

- a. Set
- b. Merge
- c. Freq
- d. Subset

iv) Which of the following program is the correct SAS statement that will subset a data set 'Test' having two variables x and y?

a. Data Test1;
Set test;
Where x = z
run;

b. data=xyz.test;
set test1;
where x =y;
Run ;

c. data test1;
set test;
Run ;

d. data test1;
set test;
where x>y;
Run ;

v) All of the following statements about Proc Freq procedure are True, except.

- a. Tables statement determines, for which variables, the frequency will be calculated.
 - b. Norow, nocol and nopercnt are valid options in Tables statement.
 - c. 'Fisher' option in tables statement will calculate Fisher's exact test
 - d. This procedure does not produce any visible output.
-