

(3 Hours)

[ Total Marks : 60

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

**Section I**

1. (a) The following data give the numbers of car thefts that occurred in a city in the past 12 days. 5

6   3   7   11   4   3   8   7   2   6   9   15

Calculate the mean, median, range, variance and standard deviation.

- (b) A group of 150 randomly selected CEOs was tested for personality type. The table gives the results of this survey : 5

	Type A	Type B
Men	78	40
Women	19	13

If one CEO is selected at random from this group, find the probability that this CEO.

- (i) has a type A personality
- (ii) is a woman
- (iii) is a man given that he has a Type A personality
- (iv) has a Type B personality given that she is a woman
- (v) has a Type A personality and is a woman
- (vi) is a man or has a Type B personality.

Are the events "Woman" and "Type A" mutually exclusive ?

What about the events "Type A" and "Type B" ?

Why or why not ?

Are the events "Type A" and "Man" independent ? Why or why not ?

2. (a) Of the volunteers donating blood in a clinic, 80% have the Rhesus factor present in their blood. If five of the volunteers are randomly selected (i) What is the probability that at least one does not have the Rhesus factor ? (ii) What is the probability that at most four have the Rhesus factor ? (iii) Find expected value and variance of the random variable. 5

- (b) Let Y have the density function given by 5

$$f(y) = \begin{cases} c(2-y), & 0 \leq y \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

- (i) Find C
- (ii) Find the mean and variance of Y
- (iii) Find  $P[1 \leq y \leq 2]$

[ TURN OVER

3. (a) The joint probability density function for the random variables  $X$  and  $Y$  which were the proportions of two components in a sample from a mixture of insecticide is given by 4

$$f(x, y) = \begin{cases} 2, & 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq x + y \leq 1 \\ 0 & \end{cases}$$

Are  $X$  and  $Y$  independent ?

- (b) Define moment generating function. Obtain MGF of Poisson distribution. 3

- (c) Let  $X$  and  $Y$  have following joint pmf. 3

y	x	
	0	1
0	0.38	0.17
1	0.14	0.02
2	0.24	0.05

Are  $X$  and  $Y$  independent ? Why or why not ?

### Section II

4. (a) Let  $X_1, X_2, \dots, X_n$  denote iid random variables from  $U(0, \theta)$  with pdf. 5

$$f(x, \theta) = \begin{cases} \frac{1}{\theta} ; & 0 < x < \theta, \theta > 0 \\ 0 ; & \text{otherwise} \end{cases}$$

Obtain the MLE and moment estimator of  $\theta$ . Find the variance of your moment estimator.

- (b) Let  $X_1, X_2, \dots, X_n$  denote a random sample from binominal  $B(n, p)$  distribution,  $n$  known. Obtain MLE of  $P$  and CRLB for the variance of unbiased estimator of  $p$ . Obtain moment estimator of  $p$ . Is it unbiased ? Find its variance. 5

5. (a) The manufacturer of a gasoline additive claims that the use of this additive increases gasoline mileage. A random sample of six cars was selected and these cars were driven for one week without the gasoline additive and then for one week with the gasoline additive. The following table gives the miles per gallon for these cars without and with the gasoline additive. 5

Without	24.6	28.3	18.9	23.7	15.4	29.5
With	26.3	31.7	18.2	25.3	18.3	30.9

- (i) Construct a 99% confidence interval for the mean  $\mu_d$  ( $d = \text{without} - \text{with}$ ).  
 (ii) Using the 2.5% significance level, can you conclude that the use of the gasoline additive increases the gasoline mileage ?  
 Assume that the population of paired differences is approximately normally distributed.



- (b) A food company is planning to market a new type of frozen yogurt. However, before marketing this yogurt, the company wants to find what percentage of the people like it. The company's management has decided that it will market this yogurt only if at least 35% of the people like it. The company's research department selected a random sample of 400 persons and asked them to taste this yogurt. Of these 400 persons, 112 said they liked it. 5

(i) Testing at the 5% significance level, can you conclude that the company should market this yogurt ?

(ii) Construct 90% confidence interval for the proportion of people(p) like yogurt.

6. (a) Violence and lack of discipline have become major problems in schools in the United States. A random sample of 300 adults was selected and they were asked if they favor giving more freedom to school teachers to punish students for violence and lack of discipline. The two way classification of the responses of these adults is presented in the following table : 6

	In favour	Against	No opinion
Men	93	70	12
Women	87	32	6

Using  $\alpha = 0.05$  test whether gender and opinions of adults are independent.

- (b) A sample of 12 observations taken from a normally distributed population produced the following data : 4

13	15	9	11	8	19
17	9	10	14	16	12

Find sample variance and construct 95% confidence interval for the population variance  $\sigma^2$ .

---

(3 Hours)

[ Total Marks : 60

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

**Section I**

1. Explain in details principles of sampling. 5
2. Design a questionnaire to conduct product testing of 2 soup flavours. 10
3. Explain differentiation of qualitative versus quantitative research. 10
4. Draw a tree diagram for syndicated services consumers. 5

**Section II**

5. (a) Why do we use Hierarchical and K-Means methods in cluster analysis ? 5
- (b) A mobile service providing company 'Cingular' has done a survey to know customers' opinion on the services offered by it. Company has used factor analysis and following is the output. What information can company extract from it : 10

**KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	.827
Bartlett's Test of Sphericity Approx. Chi-Square	1321.828
df	55
Sig.	.000

**Rotated Component Matrix<sup>a</sup>**

	Component	
	1	2
Cingular has favorable contract requirements	.775	.085
Cingular has lower prices	.730	.094
Cingular has calling plans that meet consumers' needs	.719	.048
Cingular provides high quality customer service	.668	.254
Cingular has error-free billing statement	.619	.282
Cingular has selection of phones that meet consumers' needs	.479	.270
Cingular has conveniently located stores	.432	.257
Consumer is able to make or receive calls where he/she lives and travels	.086	.850
Voice Quality of the calls is very good	.221	.768
Cingular's coverage area meets your needs	.223	.693
Consumer gets few dropped calls	.167	.682

Extraction Method : Principal Component Analysis.

Rotation Method : Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

[ TURN OVER



6. (a) How do we use factor analysis in attribute based perceptual mapping ? What is the advantage of attribute based of perceptual mapping over non-attribute based method ? 5
- (b) An IT company has done discriminant analysis where objective is to find whether its existing clients will continue to do business with it. The company has used discriminant analysis where two categories were considered viz. 'will continue' and 'will not continue'. There were 107 clients who said that they will not continue and 360 customers said they will continue. Following is the output obtained after using discriminant analysis. Interpret the output. 10

### Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
Q8. Company cares about the society	.842	87.497	1	465	.000
Q9. Highly ethical company	.846	84.885	1	465	.000
Q10. Leader in the IT industry	.757	149.425	1	465	.000
Q11. Innovative company	.729	172.898	1	465	.000
Q12. Company has strong capable senior leaders	.772	137.010	1	465	.000
Q13. Financially sound company	.882	62.309	1	465	.000
Q14. Company I can trust	.726	175.172	1	465	.000
Q15. Company has advertising I really like	.865	72.691	1	465	.000
Q16. Known for treating its employees well	.887	59.463	1	465	.000
Q17. Company does its fair share to help society	.841	87.716	1	465	.000

### Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.646 <sup>a</sup>	100.0	100.0	.626

- a. First 1 Canonical discriminant functions were used in the analysis.

### Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.608	229.113	10	.000

[ TURN OVER

### Canonical Discriminant Function Coefficients

	Function
Q. 8 Company cares about the society	1
Q. 9 Highly ethical company	-.021
Q 10. Leader in the IT industry	.053
Q 11. Innovative company	.583
Q 12. Company has strong capable senior leaders	.389
Q 13. Financially sound company	.295
Q 14. Company I can trust	-.085
Q 15. Company has advertising I really like	.511
Q 16. Known for treating its employees well	.112
Q 17. Company does its fair share to help society	-.125
(Constant)	.168
	-7.556

Unstandardized coefficients

### Functions at Group Centroids

Code	Function
	1
Will not continue	-1.471
Will continue	.437

Unstandardized canonical  
discriminant functions  
evaluated at group means

### Classification Results<sup>a</sup>

Code			Predicted Group Membership		Total
			Will not continue	Will continue	
Original	Count	Will not continue	88	19	107
		Will continue	65	295	360
		Ungrouped cases	7	8	15
	%	Will not continue	82.2	17.8	100.0
		Will continue	18.1	81.9	100.0
		Ungrouped cases	46.7	53.3	100.0

a. 82.0% of original grouped cases correctly classified.



# III Regression & Linear model

## Jan-2014.

on. 10367-13.

BB-14855

(3 Hours)

[Total Marks : 60]

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

### Section I

1. (a) The following questions are on simple Linear regression : 4  
 $y = \beta_0 + \beta_1 x + e$  where  $e \sim N(0, \sigma^2)$ 
  - (i) Prove that regression S.S. =  $\hat{\beta}_1^2 \sum (x_i - \bar{x})^2$
  - (ii) State formulae for  $\hat{\beta}_0$ ,  $\text{Var}(\hat{\beta}_0)$ ,  $\hat{\beta}_1$ ,  $\text{Var}(\hat{\beta}_1)$ ,  $\hat{\sigma}^2$
  - (iii) Suggest two tests to determine whether response variable and predictor variable are correlated.
- (b) Suppose simple linear regression analysis provides the following results.  $u = \beta_0 + \beta_1 x + e$ . 6  
 $\hat{\beta}_0 = 1$ ,  $\hat{\beta}_1 = 2$ ,  $\text{s.e.}(\hat{\beta}_0) = 0.5$ ,  $\text{s.e.}(\hat{\beta}_1) = 0.25$   
 $SST = 117.2873$ ,  $SSE = 30.0$ ,  $n = 24$ ,  $t_{22, 0.025} = 2.074$ 
  - (i) test the hypothesis  $H_0 : \beta_1 = 0$  against  $H_1 : \beta_1 \neq 0$  at 5% level of significance.
  - (ii) find value of  $y$  when  $x = 3$ .
  - (iii) find coefficient of determination.
2. (a) Multiple linear regression problem has 6 predictor variables and 30 observations. Regression and total sum of squares are respectively 3147.97 and 4296.97. Find multiple correlation coefficient. Find the value of test statistic for testing the hypothesis that all predictor variables have no explanatory power. 4
- (b) If the following assumptions do not hold, state in each case, what the problem is called as. 3
  - (i) Constancy of error variance.
  - (ii) Independence of errors.
  - (iii) Independence of predictor variables.
- (c) What is hat matrix? Express fitted values in terms of elements of hat matrix and observed responses. 3
3. (a) 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 on the effect of each variable on odds ratio.  $X_1 : (0.77, 2.51)$ ,  $X_2 : (1.23, 1.98)$ ,  $X_3 : (0.01, 0.93)$  5
- (b) Eigen values and corresponding eigen vectors of correlation matrix of three predictor variables are given below. Write principal components and their variance-covariance matrix. 3  
 $\lambda_1 = 2.23$ ,  $\lambda_2 = 1.58$ ,  $\lambda_3 = 0.19$ 

$V_1$	$V_2$	$V_3$
0.176	-0.509	.241
0.431	0.343	.406
-0.607	0.166	-.647
- (c) Describe Mallows  $C_p$  criterion. 2

[TURN OVER]

Con. 10367-BB-14855-13.

2

## Section II

4. (a) From time to time unknown to its employees the research department at post bank observes various employees for their work productivity. Recently this department wanted to check whether the four tellers at a branch of this bank serve, on average the same number of customer per hour. The research manager observed each of the four tellers for a certain number of hours. The following table gives the number of customers served by the four tellers during each of the observed hours.

teller A	teller B	teller C	teller D
19	14	11	24
21	16	14	19
26	14	21	21
24	13	13	26
18	17	16	20
	13	18	

At the 5% significance level, test the null hypothesis that mean number of customers served per hour by each of these four tellers is the same. Find the 95% confidence interval between the difference of teller B and D. Given : Total sums of squares = 413.82.

- (b) Explain the difference between t-test and one way model.
5. (a) Explain the Q-Q plot. Construct the Q-Q plot from the following data :—  
 $-3, -6, -4, 2, 5, 3, 1, 6, 4, 2.6, 1.5, 3.2$
- (b) Four experimenters determine the moisture content of samples of a powder, each man taking a sample from each of six consignments. The assessments are :

Observer	Consignment					
	1	2	3	4	5	6
1	9	10	9	10	11	11
2	12	11	9	11	10	10
3	11	10	10	12	11	10
4	12	13	11	14	12	10

Carry out the ANOVA and discuss whether there is any significant difference between consignments or between observers. Use the following information  $y.. = 259$   
 Total sum of squares = 35.9583

- 6 (a) Explain nested model with example.
- (b) Analyse the following data from a two way nested classification.

Course	Section of courses	individuals
English	1	2, 3, 4
	2	8, 10, 9, 5
Geology	1	6, 8, 10, 2
	2	6, 2, 1
	3	2, 4
	4	8, 6



(3 Hours)

[ Total Marks : 60

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

**Section I**

1. (a) Explain the relative strengths and weaknesses of the Regression Model and the Time Series Model. **10**  
 (b) A fifty-year old pharmaceutical company wants quarterly forecasts of its sales in the next two years. Present a suitable action plan for this purpose.
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** : Forecasting future values of a cardinal variable is possible even if past quantitative data on it are not available.  
**Statement Number 2.2** : Moving averages of period 3 (three) should effectively smoothen the quarterly sales figures of cold drinks.  
**Statement Number 2.3** : Strong correlation between two variables does not necessarily imply a cause and effect relationship between the two.  
**Statement Number 2.4** : It is not necessary to compute correlation matrix before starting work on a multivariate regression model.  
**Statement Number 2.5** : A detailed error analysis must be carried out before confirming goodness of a forecasting model.
3. (a) Describe the process of 'Six Thinking Hats Technique' of decision making and explain its rationale. **10**  
 (b) How will you use it to decide whether a two-wheeler manufacturing company should diversify into computer software business ?

**Section II**

4. Write short notes on any **five** :—

- (a) Demand
- (b) Lead time
- (c) Order Cycle
- (d) Re-order level
- (e) Stock replenishment
- (f) Inventory turnover
- (g) Standardisation.

**10****[ TURN OVER**

5. Define problem of Economic Order Quantity with known demand. The shortages are not allowed. Rate of replenishment of inventory is instantaneous. Determine optimum production quantity for each production run of unequal length stating the conditions and diagram. **10**

### Section III

6. Shift Ltd. has been offered a choice to buy a machine between i10 and i20 make. **10**  
You are required to compute :—
- Breakeven point for each of the machines
  - The level of sales at which both machines earn equal profits
  - The range of sales at which one is more profitable than the other.

The other relevant data is as given below :—

Particulars	i10 make	i20 make
Annual output in Units	10,000	10,000
Fixed Cost	30,000	16,000
Profit at above level of production	30,000	24,000

The Market price of the final product is expected to be ₹ 10 per unit.



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

## SECTION I

Instructions:

1. Use of Statistical Software & Excel along with six sigma excel template is allowed
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. FMEA  
c. DMAIC Phases of 6 Sigma  
b. 5 S, Kaizen  
d. Gage R&R

(10)

**Q2.**

- a) Which IPL Cricket Team is better & reliable w.r.t. total score made in T20 Series.

(10)

Mumbai Indians	155	145	140	134	149	150	162	157	139	142	156	143
CSK	145	160	132	154	159	160	152	168	148	140	162	153

- b) Given below M/C 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
Efficiency %	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 Viscosity of product A. Conduct the DOE with replication & write down the regression equation based on p-value.

(10)

F1 : Residence Time

F2 : Water Application Rate Kgs/Hr mm))

F3 : Steam Flow

F4 : Cylinder Temp.

Run	F1	F2	F3	F4	Ceramic Strength	Random Order	Run	F1	F2	F3	F4	Ceramic Strength	Random Order
1	-1	-1	-1	-1	530.45	17	17	-1	-1	-1	-1	457.34	12
2	1	-1	-1	-1	572.48	30	18	1	-1	-1	-1	470.80	1
3	-1	1	-1	-1	552.14	14	19	-1	1	-1	-1	460.55	4
4	1	1	-1	-1	516.93	8	20	1	1	-1	-1	488.04	23
5	-1	-1	1	-1	553.67	32	21	-1	-1	1	-1	435.19	2
6	1	-1	1	-1	492.14	20	22	1	-1	1	-1	436.17	28
7	-1	1	1	-1	542.98	26	23	-1	1	1	-1	451.67	11
8	1	1	1	-1	519.26	24	24	1	1	1	-1	458.31	9
9	-1	-1	-1	1	341.58	10	25	-1	-1	-1	1	292.90	25
10	1	-1	-1	1	325.52	16	26	1	-1	-1	1	284.41	21
11	-1	1	-1	1	328.76	27	27	-1	1	-1	1	267.66	6
12	1	1	-1	1	418.23	18	28	1	1	-1	1	360.84	7
13	-1	-1	1	1	294.72	3	29	-1	-1	1	1	242.11	5
14	1	-1	1	1	260.37	19	30	1	-1	1	1	193.22	13
15	-1	1	1	1	278.51	31	31	-1	1	1	1	235.52	22
16	1	1	1	1	341.47	15	32	1	1	1	1	296.73	29

[ TURN OVER



## SECTION II

(05)

4.(a)  
A manufacturing company outsourced its after sales service activity to another company. They wanted to get a feel of type of complaints and its volume in the last 6 months and which type of complaints are maximum and need immediate attention. They asked the vendor to submit graphical representation of data, which can be presented to the management. Which quality tool you think the vendor should use to represent the data? Draw appropriate diagram to represent the data given below.

Complaint Type	Number of complaints
A	27
B	95
C	6
D	15
E	46
F	7

4.(b)

(05)

Suggest appropriate QC tool from 7 new QC tools that can be used for following problems.

1	A washing powder has different efficiencies at achieving 'softness' and 'stain removal' in garments made of acrylic, polyester, wool and various fiber mixtures. If similar affects are found in a group of fibers, then changing the powder ingredients may affect the whole group in a similar way.
2	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.
3	A class of students have various ideas about how to make learning more interesting. We need to group them and arrive at the most effective ideas.
4	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.
5	To calculate the earliest date the project can be completed, and to find ways of changing this

[ TURN OVER



4.c.

(05)

Students have to do a project as part of their final semester. A group of 6 students decided to visit a manufacturing unit and collect their quality data for analysis and use it for the project. It involved following activities.

Srl	Activity	Time required in Days
1	Permission of the managing director of the manufacturing unit	8
2	Booking of ticket for travel	10
3	Collection of data at the unit	2
4	Analysis of data set 1 with 2 students	3
5	Analysis of data set 2 with 2 students	5
6	Analysis of dataset 3 with 2 students	4
7	Final analysis based on analysis of data set 1, 2 and 3	2
8	Writing of project Report	3
9	Printing 3 copies of the report by professional printer	2
10	Submission of the project	1

1. Draw activity diagram for these activities.
2. Find our critical path
3. If the project submission date is 1<sup>st</sup> July then when should travel tickets be booked latest?

5.a.

(05)

Meteorological department wanted to find out whether there is any relationship between amount of rain fall during day on rainy day and minimum temperature at night on that day during rainy season. They collected data as follows. Which QC tool can be used for this?

What is your conclusion based on this data?

Rain fall in mm	Min Temperature in degree cc
2.2	27
3.2	26.5
1.2	29
0.8	30.5
3.6	25.5
4.8	22
2.6	26
4.5	21
2.4	25
0.3	28
0.6	29
3.5	26
4.4	23.5
0.5	29
1.8	27

[ TURN OVER

5.b.

(05)

Draw Control Chart for following data and suggest how many items are outside limit

- Average weight = 10 gm
- Upper Limit = + 0.20 gm
- Lower Limit = - 0.20 gm

Weight of sample items is as follows.

10.05, 10.10, 10.10, 9.95, 9.90, 10, 9.75, 9.95, 10, 10.05

5.c.

(05)

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.

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)



- 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 tests/methods in analysis of clinical research data. [ Any 4 of 5, 2.5 Marks each ] (10)

- ANCOVA
- Two sample t test
- Logistic Regression
- Chi Square Test
- Odds Ratio

2. Describe contents of statistical analysis plan for a clinical trial project. (10)

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

- Explain Type I error, type II error and power of a test
- Explain purpose of Randomization and Blinding in a clinical Trial

4. What are ICH guidelines? Briefly discuss the purpose of E6, E9 and E3 guideline (10)

5. Please read the situation described below and answer the questions mentioned below. (10)

Efficacy, safety, and tolerability of oral Mycophenolate Mofetil plus Corticosteroids were compared with that of Intravenous Cyclophosphamide plus corticosteroids for treatment of active lupus nephritis in an open-label, randomized study.

The primary end point was achievement of complete remission at 24 weeks (criteria to decide whether remission is achieved or not, are described in protocol. Achievement of remission – or otherwise – will be determined by treating investigator based upon these criteria).

The protocol was designed as a superiority trial to demonstrate that Mycophenolate Mofetil is superior to intravenous cyclophosphamide as induction therapy to produce remission in active lupus nephritis.

- a. Describe the null and alternative hypothesis for this study?
- b. What information will you need to calculate sample size for this study?
- c. Please discuss statistical methodology to test hypothesis in this study

[ TURN OVER

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

**GENERAL INSTRUCTIONS TO ANSWER QUESTIONS BY SELECTING CHOICES**

To answer a question, mention Question number followed by the alphabet representing the correct choice in your answer paper.

Example:

1. Sample question text: \_\_\_\_\_?

- |             |             |
|-------------|-------------|
| a. Choice 1 | b. Choice 2 |
| c. Choice 3 | d. Choice 4 |

If your answer is "Choice 4", then write following in your Answer paper.

Q. 1: d

- 1) All of the following statements about Proc Power procedure are True, except?
  - a. It can be used for calculating Sample size or Power of a study.
  - b. It cannot be used to calculate sample size for Exact Tests.
  - c. It can generate a graphical output
  - d. It can be used to provide sample size for one sided as well as two sided test.
- 2) Which of the following statement is correct?
  - a. Proc Sort procedure cannot create a new SAS dataset.
  - b. Proc SQL cannot retrieve distinct values of variables in a SAS dataset.
  - c. Proc Plan procedure can be used to generate 1:1 randomization schedule
  - d. Proc Format procedure is used to produce formatted RTF output.
- 3) PROC FREQ procedure is commonly used for all of the following except
  - a. To produce inferential statistics for continuous data.
  - b. To produce confidence interval for Odds ratio.
  - c. To generate exact binomial confidence interval for success proportion in single sample
  - d. To generate Fisher's Exact Test statistic.

[ TURN OVER



4) Which of the following statement about the code below is correct?

```
Proc SQL;
```

```
Select * From Work.ABCD;
```

```
QUIT;
```

- a. Summary of All continuous variables from ABCD dataset in work library will be displayed.
- b. ABCD dataset will be created in work library.
- c. Rows from ABCD dataset in work library will be displayed and will be ordered based upon first variable on the dataset
- d. All rows from ABCD dataset will be displayed.

5) Identify the missing.

```
Data exam;
```

```
Number=roll/10;
```

```
Run;
```

```
Proc sort data=exam;
```

```
_____ number;
```

```
Run;
```

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

\*\*\*\*\*

**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**.

1. (a) Describe how statistical data having multivariate observations can be summarized using (i) Descriptive statistics (ii) Graphical method. 4
- (b) Consider a random vector  $X = [X_1 \ X_2 \ X_3 \ X_4]^T$  with mean  $\mu = [4 \ 2 \ 3 \ 1]^T$  and variance covariance matrix. 6

$$\Sigma = \begin{bmatrix} 4 & 0 & 3 & 0 \\ 0 & 1 & 0 & -2 \\ 3 & 0 & 9 & 0 \\ 0 & -2 & 0 & 4 \end{bmatrix}$$

If  $A = [2 \ 1 \ 3 \ 1]$  obtain

- (i) Mean and variance of  $AX$
- (ii) Correlation coefficient between (1)  $X_1$  and  $X_3$  and (2)  $X_2$  and  $X_3$ .
- (iii) Covariance matrix between  $Y_1$  and  $Y_2$  where  $Y_1 = [X_1 \ X_3]^T$  and  $Y_2 = [X_2 \ X_4]^T$ . Hence comment.

2. (a) Let  $X_1 = \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}$   $X_2 = \begin{bmatrix} 3 \\ 2 \\ 4 \end{bmatrix}$   $X_3 = \begin{bmatrix} 4 \\ 3 \\ 0 \end{bmatrix}$  be a random sample from a joint distribution 3

that has mean vector  $\mu$  and covariance matrix  $\Sigma$ . Obtain the unbiased estimator of  $\mu$  and  $\Sigma$ .

- (b) The p.d.f. of a 4-variate normal distribution is 3

$$\frac{1}{(2\pi)^2(30)} \exp \left[ -\frac{1}{2} \left\{ \frac{(x_1-3)^2}{5} + \frac{(x_2-4)^2}{3} + \frac{x_3^2}{10} + \frac{(x_4-1)^2}{6} \right\} \right]$$

Obtain mean vector and variance covariance matrix.

- (c) Explain the method of  $\chi^2$  plot to test normality of the given data. 4

[ TURN OVER



3. (a) Let  $X \sim N_p(\mu, \Sigma)$ ,  $|\Sigma| > 0$ . 5

(i) State distribution of  $AX$  where  $A$  is a  $5 \times p$  matrix of constants.

(ii) State the conditional distribution of  $X_{(1)}$  given  $X_{(2)} = x_{(2)}$  when

$$X = \begin{bmatrix} X_{(1)} \\ X_{(2)} \end{bmatrix}, \mu = \begin{bmatrix} \mu_{(1)} \\ \mu_{(2)} \end{bmatrix} \text{ and } \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Sigma_{22} > 0.$$

(iii) State distribution of  $(X - \mu)' \Sigma^{-1} (X - \mu)$ .

(b) Let  $\underline{X} \sim N_p(\mu, \Sigma)$ . State MLE of  $\mu$ ,  $\Sigma$  and  $\mu' \Sigma^{-1} \mu$ . 2

(c) Write a **short** note on 'Transformations to Near Normality'. 3

4. (a) A random sample of size 4 from a bivariate normal population gave the following :— 4

$$\bar{X} = \begin{bmatrix} 9 \\ 6 \end{bmatrix} \quad S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$$

Use Hotelling's  $T^2$  statistic to test the hypothesis  $H_0 : \mu = [10 \ 5]^1$  (Use  $F_{22}(5\%) = 19.00$ )

(b) Explain the use of  $T^2$  statistic for testing the equality of mean vectors from two independent multivariate normal populations with same covariance matrix  $\Sigma$ . 4

(c) **State** the multivariate linear regression model. 2

5. (a) Consider the following distance matrix. 10

$$\begin{array}{c} 1 \quad 2 \quad 3 \quad 4 \quad 5 \\ \begin{bmatrix} 0 & & & & \\ 9 & 0 & & & \\ 3 & 7 & 0 & & \\ 6 & 5 & 9 & 0 & \\ 11 & 10 & 2 & 8 & 0 \end{bmatrix} \end{array}$$

Cluster the five items using complete linkage. Draw the dendrogram.

[ TURN OVER

(b) Let  $X = [X_1, X_2]^T$

$$\Sigma = V(X) = \begin{bmatrix} 1 & 4 \\ 4 & 100 \end{bmatrix}.$$

Obtain (i) the principal components based on  $\Sigma$

(ii) principal components based on correlation matrix and comment.

6. (a) Explain the principal component method to extract the initial factor loading. **10**  
Write down the orthogonal factor model. Explain the terms communalities specific variances.

(b) Define the probabilities of misclassification in a classification problem. State min. ECM rule.

For the two populations  $\pi_1$  and  $\pi_2$  based on a sample of size three, sample mean vectors and sample pooled covariance matrix given below :—

$$\bar{x}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix} \quad \bar{x}_2 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad S_{\text{pooled}} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

Calculate linear discriminant function.

-----