

( 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-books.

**Section I**

1. (a) The following data give the numbers of driving citations received by 12 drivers :— 5

4 8 0 3 11 7 4 14 8 13 7 9

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

- (b) A random sample of 250 adults was taken, and they were asked whether they prefer watching sports or opera on television. The following table gives the two-way classification of these adults :— 5

	Prefers Watching Sports	Prefers Watching Opera
Male	96	24
Female	45	85

If one adult is selected at random from this group, find the probability that this adult :—

- Prefers watching opera
- is a male
- Prefers watching sports given that the adult is a female.
- is a male given that he prefers watching sports.
- is a female and prefers watching opera.
- Prefers watching sports or is a male.

Are the events "female" and "prefers watching sports" independent ? Are they mutually exclusive ? Explain why or why not.

2. (a) Let X be a random variable with a density function given by :— 5

$$f(x) = \begin{cases} C x^2 & ; -1 \leq x \leq 1 \\ 0 & ; \text{otherwise} \end{cases}$$

Find the value of x that makes f(x) a probability density function. Find its mean, variance and median.

- (b) On average, 5.4 shoplifting incidents occur per week at an electronics store. Find the probability that exactly three such incidents will occur during a given week at this store. 2

[ TURN OVER

- (c) According to U.S. Census Bureau data, 41% of married couples have no children. Assume that this result is true for current population of all U.S. married couples. What is the probability that in a random sample of 200 married couples, 90 or more have no children ? 3

3. (a) Let  $X$  and  $Y$  be random variables with density function :- 5

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

Find the marginal pdf of  $x$  and  $y$ -also find conditional pdf of  $y$  given  $x$  and  $E(y|x)$ .

- (b) Suppose the random variable  $X$  and  $Y$  have following joint pmf 5

X			
Y	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

Find marginal pmf of  $X$  and  $Y$ .

Find conditional pmf of  $X|Y = 2$ .

Find covariance between  $X$  and  $Y$ .

### Section II

4. (a) Let  $X_1, X_2, \dots, X_n$  demote a random sample from the normal distribution with mean  $\mu$  and variance 1. Obtain MLE and moment estimator of  $\mu$ . Is it unbiased ? Find its variance. 5
- (b) Let  $X_1, X_2, \dots, X_n$  be a random sample from Poisson  $P(\lambda)$  distribution. Obtain MLE and moment estimator of  $\lambda$ . Find CRLB. Find expectation and variance of your estimate. 5

5. (a) Many students suffer from math anxiety. A professor who teaches statistics offered her students a 2-hour lecture on math anxiety and ways to overcome it. The following table gives the test scores in statistics of seven students before and after they attended this lecture : 5

Before	56	69	48	74	65	71	60
After	62	73	44	85	71	70	73

- (i) Construct a 99% confidence interval for the mean  $\mu_d$  ( $d = \text{Before} - \text{After}$ ).  
 (ii) Test at the 2.5% significance level whether attending this lecture increases the average score in statistics.

Assume that the population of paired differences is approximately normally distributed.



- (b) A consumer agency wants to estimate the proportion of all drivers who wear seat belts while driving. A random sample of 150 drivers has shown that 76% of drivers wear seat belts while driving. Construct 95% confidence interval for the population proportion. 5

6. (a) The U.S. Bureau of Labor statistics compiles data on the employment status of high school dropouts aged 18-24. The following table classifies such individuals by gender and employment status as of October 1997. Note that the numbers in the tables are in thousands. In the table, Not in Labor Force means that these individuals are neither working nor looking for a job :- 6

	Employed	Unemployed	Not in Labor Force
Men	311	52	497
Women	440	49	506

Suppose that this table was based on a random sample of high school dropouts. At the 5% level of significance, can you conclude that employment status and gender are related for all such dropouts ?

- (b) An auto manufacturing company wants to estimate the variance of miles per gallon for its auto model AST727. A random sample of 22 cars of this model showed that the variance of miles per gallon for these cars is 0.62. Construct the 95% confidence interval for the population variance. 4

\*\*\*\*\*

(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 detail different types of measurement scales. 5
2. Design a questionnaire for launching a new mobile brand in the market. 10
3. Explain differentiation in focus groups, depth interviews and projective techniques. 10
4. Draw a tree diagram for classification of MR data. 5

**Section II**

5. (a) Why do we use Hierarchical and K-Means methods in cluster analysis ? 5
- (b) A Bank having branches in different countries has done a survey to know customers' opinion on the services offered by the bank. Bank has used factor analysis and following is the table of rotated component matrix. What information can Bank extract from it ? 10

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

	Component		
	1	2	3
Q6_j. Representative knowing your overall situation and needs	.911	.170	.103
Q6_l. Degree to which my provider knows me	.878	.088	.101
Q6_h. Quality of advice	.839	.232	.101
Q6_i. Knowledge of representatives or advisors you deal with	.811	.266	.166
Q6_e. Ability to resolve problems	.566	.364	.348
Q6_c. Fees or commissions charged	.008	.812	.186
Q6_b. Performance of investments with this provider	.333	.738	-.009
Q6_d. Depth of products and services to meet the range of your investment needs	.262	.636	.461
Q6_m. Quality of Service	.541	.584	.164
Q6_f. Online services offered	-.048	.114	.858
Q6_g. Multiple providers' products to choose from	.246	.231	.752
Q6_k. Access to other professional resources	.527	.041	.597

Extraction Method : Principal Component Analysis. Rotation Method : Varimax with Kaiser Normalization. a. Rotation converged in 5 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) Bank has done discriminant analysis where objective is to find relation between whether customer will recommend the bank to others ? Bank has used discriminant analysis where two categories were considered viz. 'will recommend' and 'will not recommend'. There were 40 customers who said that they will not recommend and 324 customers said they will recommend. Following is the output obtained after using discriminant analysis. Interpret the output. 10

#### Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.483 <sup>a</sup>	100.0	100.0	.571

a. First 1 canonical discriminant functions were used in the analysis.

#### Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	sig.
Q6_b. Performance of investments with this provider	.812	83.594	1	362	.000
Q6_c. Fees or commissions charged	.927	28.302	1	362	.000
Q6_d. Depth of products and services to meet the range of your investment needs	.876	51.138	1	362	.000
Q6_e. Ability to resolve problems	.884	47.361	1	362	.000
Q6_f. Online services offered	.981	7.018	1	362	.008
Q6_g. Multiple providers' products to choose from	.937	24.255	1	362	.000
Q6_h. Quality of advice	.891	44.125	1	362	.000
Q6_i. Knowledge of representatives or advisors you deal with	.886	46.522	1	362	.000
Q6_j. Representative knowing your overall situation and needs	.918	32.173	1	362	.000
Q6_k. Access to other professional resources	.943	21.766	1	362	.000
Q6_l. Degree to which my provider knows me	.929	27.541	1	362	.000
Q6_m. Quality of service	.709	148.474	1	362	.000

**Canonical Discriminant Function  
Coefficients**

	Function
	1
Q6_b. Performance of investments with this provider	.271
Q.6_c. Fees or commissions charged	.005
Q.6_d. Depth of products and services to meet the range of your investments needs	.029
Q.6_e. Ability to resolve problems	.079
Q.6_f. Online services offered	-.015
Q.6_g. Multiple providers' products to choose from	.15
Q.6_h. Quality of advice	.017
Q.6_i. Knowledge of representatives or advisors you deal with	.100
Q.6_j. Representative knowing your overall situation and needs	-.151
Q.6_k. Access to other professional resources	.034
Q.6_l. Degree to which my provider knows me	-.040
Q.6_m. Quality of service (Constant)	-4.519

Unstandardized coefficients

**Functions at Group  
Centroids**

code	Function
	1
.00	-1.972
1.00	.243

Unstandardized canonical  
discriminant functions  
evaluated at group means

-----



- 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. Assume regression model  $Y = \beta_0 + \beta_1 x + e$

- (a) Use the result : Regression ss =  $\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$  to estimate  $\beta_0, \beta_1$ . Also determine  $R^2$  and  $\hat{\sigma}^2$  using the following information : 5

$R^2$  and  $\hat{\sigma}^2$  using the following information :

$$\sum (y_i - \hat{y}_i)^2 = 3.267, \sum y_i = 63.3, \sum y_i^2 = 423.49, \\ \sum x_i = 139, \sum x_i^2 = 2239, n = 10$$

- (b) (i) Suggest two tests to determine whether response variable and predictor variable are correlated. 5  
(ii) State formulae for s.e.  $(\hat{\beta}_0)$  and  $\text{var}(\hat{\beta}_1)$ .  
(iii) What is the sign of product  $\hat{\beta} \times \text{cov}(y, x)$  ? Explain.

2. (a) Multiple linear regression is studied using 25 observations. Give procedure of testing following hypotheses in the model. 5

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_5 x_5 + e$$

- (i)  $\beta_1 = \beta_2 = \dots = \beta_5 = 0$   
(ii)  $\beta_1 = \beta_4 = 0$  given  $\beta_2 = \beta_3 = \beta_5 = 0$

(b) Computer output of logistic regression is given below. Variable X denotes age in years and Y denotes presence or absence of coronary heart disease (CHD), where  $Y = 1$  if CMD is present and  $Y = 0$  otherwise 5

Variable	Coefft	S.e.e	Z	p value
Age	0.111	0.0241	4.61	< 0.001
Constant	- 5.309	1.1337	- 4.68	< 0.001

Estimate  $\pi(x)$  when  $x = 26, 55$  and  $75$ . How would you classify following observations ?  
What is the percent of correct classification ?

Age	CHD
X	Y
26	0
55	0
75	1

**Con. 14046-14.**

[ TURN OVER

3. (a) How principal component analysis helps to detect multicollinearity? 5  
 If the eigen values of correlation matrix of three predictor variables are  $\lambda_1 = 2.952$ ,  $\lambda_2 = 0.040$ ,  $\lambda_3 = 0.008$ . Use two methods to check presence of multicollinearity.
- (b) How do you use forward selection procedure to select variables for linear regression? 2
- (c) Why is transformation of variables necessary? State the transformations for the following models :- 3

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

### Section - II

4. (a) Explain one-way model with examples. Write down its ANOVA. 4  
 (b) Fifteen fourth grade students were randomly assigned to three groups in order to experiment with different methods of teaching arithmetic. At the end of semester the same test was given to all 15 students. The table gives the scores of students in the three groups. 6

Method I	Method II	Method III
48	55	84
73	85	68
51	70	95
65		74
87		

**Given :** Total sums of squares = 2440. Analyse the data and draw the conclusion.

5. (a) A car manufacturer is studying three devices for reducing fuel consumption using three different models of the car to be fitted with the devices. The data obtained are as follows :- 6

Device Type	Car models											
	I				II				III			
1	44	30	40	55	34	40	35	33	32	30	38	35
2	40	48	49	50	46	49	45	46	40	48	45	45
3	48	40	48	45	84	30	30	39	36	30	34	32

Analyse the data as a full two factor factorial design and test for significance of all effects at 1% level of significance.

**Given :** Total sums of squares = 1786.75

Interaction sums of squares = 162.83.

- (b) Explain nested models with examples. 4
6. (a) Explain Q - Q plots and construct the Q - Q plot from the following data :- 5  
 $-1.5, -2.5, 1, 3, 2, 2.3, -1.3, -3, -2, 0.5$ .
- (b) Explain the analysis of covariance model with examples. 5



- 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) Describe in detail the systematic process of multiple regression analysis. 10  
(b) How will you proceed to prepare a reliable forecast of demand for motorbikes in India ?
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** : Time available for forecasting is an important factor in choosing forecasting methods.  
**Statement Number 2.2** : Mean square error within acceptable limits is the sufficient proof of goodness of the forecasting model.  
**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** : The classical decomposition model of time series is highly recommended for long term forecasting.  
**Statement Number 2.5** : Absence of multi-collinearity must be confirmed before formulating a multiple regression model.
3. (a) What is the difference between Wald's Maximin Criterion and Hurwicz's Maximax Criterion in the context of structured decision making ? Illustrate the difference with a suitable decision making situation. 10  
(b) What is the rationale underlying Savage's Minimax Regret Criterion of decision making ? Explain this criterion using the same illustration used in Q. No. 3 (a) above.

**Section II**

4. What are the five different types of inventories ? Explain each of them in brief. 10
5. What are selective inventory control techniques ? Explain three main selective inventory control techniques. 10

## Section III

6. Jaya Auto Ltd. manufactures and sells three products to the automobile industry. All the products must pass through a machining process, the capacity of which is limited to 20,000 hours per annum both by equipment design and government regulation. 10

The following additional information is available :—

Particulars	Make 21	Make 31	Make 41
Selling Price ₹/unit	1,900	2,400	4,000
Variable Cost ₹/unit	700	1,200	2,800
Machining requirement Hrs/units	3	2	1
Maximum Possible sales-units	10,000	2,000	1,000

Prepare a Statement showing the best possible production mix, which would provide the maximum profit to Jaya Auto Ltd. together with supporting workings.



# Paper V Six Sigma and Statistical Process Control.

QP Code : BB-20548

(3 Hours)

July 2014.

[Total Marks:60]

- N.B. (1) All question 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 I

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.
1. Below data is in no of days taken by an Software Company for implementing (10) ERP at Client's site.. Kindly comment on their performance based on your analysis. Which year Company performed consistently month over month? What improvement Target can be set for the year 2014-15 & for consistency?

Year	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar
2012-13	220	236	222	230	233	245	229	246	235	225	245	246
2013-14	211	226	198	220	225	226	218	224	214	206	228	219

2. Data on quality index of 3 brands measured by 3 quality inspectors' is given (05) below & kindly advise Company on measurement analysis.

Brand Code	1	1	1	1	1	1	2	2	2
Inspector No	1	2	3	1	2	3	1	2	3
Trial No	1	1	1	2	2	2	1	1	1
Quality Index	215	182	229	219	202	203	109	66	42
Brand Code	2	2	2	3	3	3	3	3	3
Inspector No	1	2	3	1	2	3	1	2	3
Trial No	2	2	2	1	1	1	2	2	2
Quality Index	108	68	58	138	145	150	155	142	161

3. Please advise on below to set the base-line (05)  
a. Sample Size to be used with SD of 12 & tolerance of 3 is allowed at 95% confidence.

Con. 14039-14.

[TURN OVER



- b. What will be DPMO & Zst for having 4 defects out of 50 Units with OFE of 4 ?  
 c. What will be DPMO & Zst for USL-1050 ml & LSL of 970 ml with current variance of 100 ml?

4. Following 3 factors determine the Efficiency % on a machine. Conduct the DOE & identify the factors affecting Production. (10)

low	In-house Trained	Local	-1
high	OEM trained	imported	+1
Speed	Technical Skills	Material Quality	Efficiency %
-1	-1	-1	70.35
1	-1	-1	70.51
-1	1	-1	75.02
1	1	-1	75.21
-1	-1	1	70.43
1	-1	1	70.59
-1	1	1	75.11
1	1	1	75.31

## SECTION II

5. (a) Suggest appropriate QC tool that can be used for following problems (05)

Sr.	Problem
1	University is facing an issue of poor attendance in class and decides to involve professors from various departments to find out root cause of this and implement new policy which will benefit students in their learning.
2	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.
3	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.
4	A firm of consulting engineers wants to ensure that all eventualities and their relations are covered in an investigation report into the laying of a new cross-country gas pipeline.
5	A pharmaceutical company examines the pain-killing drugs of its subsidiaries in terms of the cost to product and general efficacy. Products which are high cost but are not of highest efficacy are dropped. Low-cost drugs of reasonable efficacy are promoted, and high-cost drugs have a project initiated to reduce production cost.

Con. 14039-14.

[TURN OVER



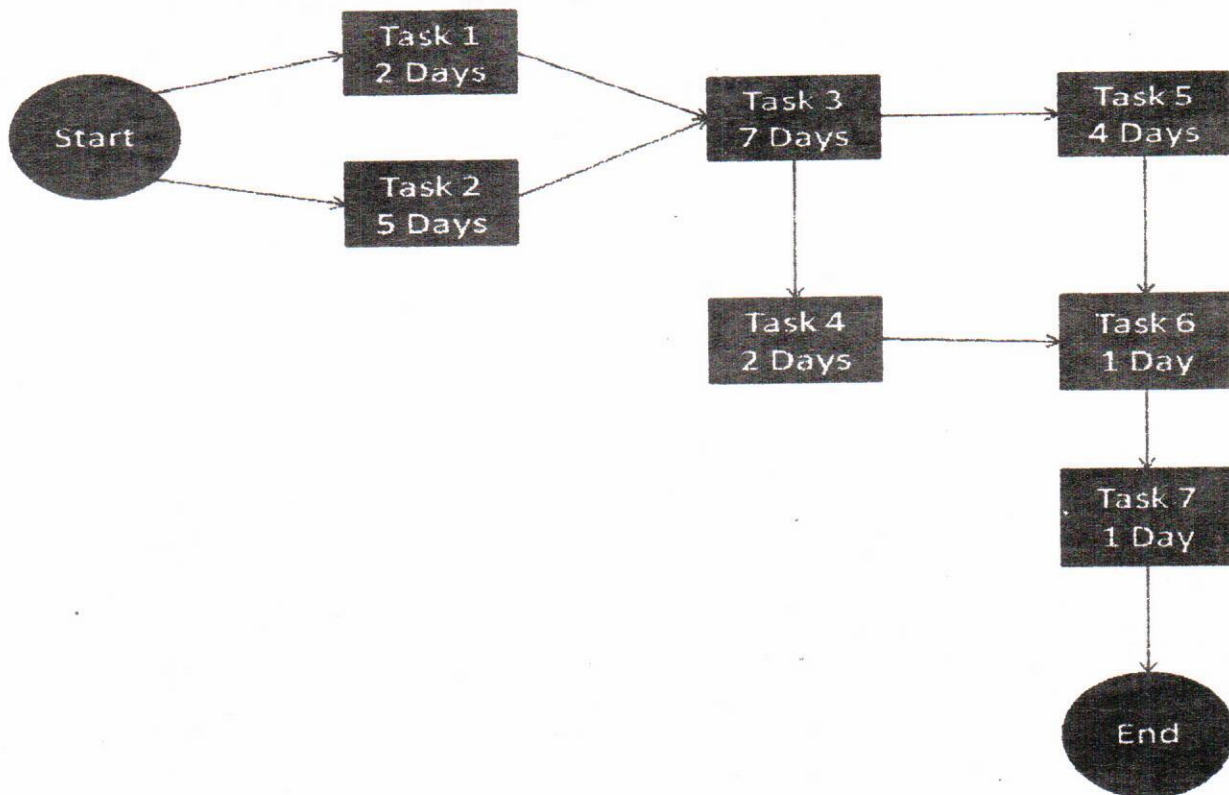
(b) Refer following excel and answer following questions

(05)

	A	B	C	D	E	F	G	H
1								
2	A1	F	57	96	48	86	39	20
3	A2	M	66	57	31	58	93	30
4	A3	M	55	47	97	29	68	28
5	A4	F	79	68	45	54	87	95
6	A5	F	30	38	60	65	87	60
7	A6	M	42	50	56	37	56	81
8	A7	M	98	37	53	31	44	54
9	A8	M	20	97	45	39	77	28
10	A9	F	46	44	73	24	23	34
11	A10	F	73	78	75	32	98	80
12	A11	M	21	44	41	87	45	92
13	A12	M	28	61	48	96	61	84
14	A13	F	98	20	71	68	59	31
15	A14	F	42	80	55	39	47	66
16	A15	M	40	85	59	77	85	93
17	A16	M	55	99	54	99	87	92
18	A17	F	34	63	81	86	31	92
19	A18	F	80	75	34	45	88	84
20	A19	M	34	44	20	89	85	93
21	A20	M	49	47	42	85	69	92

1. Write formulae to find out Average marks of Male students
2. Write formulae to calculate total marks in Maths by Male students
3. Write formulae to determine which subject out of Maths and Stats is more scoring overall (Output should be Maths or Stats)
4. Write formulae to calculate highest marks in Computers
5. Write formulae to find out, average marks in English

6. (a) Following activity diagram represents project plan of road repairs to be done before monsoon. Answer following questions. (05)



- 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 31<sup>st</sup> May?
- When can we start Task 4 (latest) without impacting overall schedule?

(b) Cold drink manufacturing company is assessing risks involved in it's launch (05) of new product, which involves significant investment and would like to ensure that top risks are dealt with appropriate counter measures that will minimise the loss in case of risk occurrence.

- Following are the risks identified with it's corresponding probability of risk occurrence. Calculate risk exposure and find out the most serious risk.



Risk	Probability of occurring	Loss if risk occurs
Product recall situation	3%	20,000
Significant product rejection	0.1%	1,000,000
Comprehensive strike	8%	25,000

- ii. Following are the countermeasure against the most serious risk with it's corresponding costs and new probabilities of reduced risks. Find out new risk exposure and risk reduction leverage.

Countermeasure	Total Cost	New Risk Probability	New Total Loss
Advertising Campaign	40,000	3%	5,000
Price Promotions	30,000	5%	10,000
Simultaneous Launch	10,000	8%	15,000

- iii. Find out the cost effective countermeasure

7. (a) Draw Control Chart for following data and suggest how many items are outside limit (05)

- 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

- (b) A case study has two 2-level (A & E) and three 3-level factors (B, C, D). (05)

Calculate degrees of freedom. Refer following table of standard orthogonal arrays and determine which orthogonal array is suitable for this case study? What technique can be used to reduce number of experiments to 9? Explain your answer and draw a table of experiment layout.



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. 14039-14.

[TURN OVER



L9 -

	A	B	C	D
1	A1	B1	C1	D1
2	A1	B2	C2	D2
3	A1	B3	C3	D3
4	A2	B1	C2	D3
5	A2	B2	C3	D1
6	A2	B3	C1	D2
7	A3	B1	C3	D2
8	A3	B2	C1	D3
9	A3	B3	C2	D1

\*\*\*\*\*

Con. 14039-14.

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

1. Describe application of following tests/statistical methods in clinical research studies (any 4 of 5 – 2.5 marks each) (10)
  - a. Wilcoxon signed rank test
  - b. Binomial Test
  - c. ANCOVA
  - d. Relative risk
  - e. Bonferroni Adjustment
2. Describe the role of a statistical Analyst in a single clinical trial project. (10)
3. Answer following questions (5 marks each) (10)
  - a. Describe objectives of clinical trials conducted in different phases of clinical drug development?
  - b. Briefly describe steps involved in clinical data management activities for a clinical trial project.
4. Briefly explain the purpose of ICH E3, ICH E6 and ICH E9 guidelines. What key considerations for Statistical analysis of a clinical trial are mentioned in ICH E9 guidance? (10)
5. Please read the following scenario and answer questions mentioned below. (10)
 

Clinical trial of a new formulation of Insulin is being planned. Insulin is used in treatment of diabetes. Sponsor of this study feels that the new formulation may not control diabetes as well as standard formulation, but is much safer than existing formulation. Sponsor also feels that loss in efficacy (beneficial effect) with new formulation is negligible as compared to existing formulation.

Therefore sponsor wants to demonstrate that the efficacy of new formulation is not worse than standard formulation by more than a certain margin. Clinicians associated with study opined that efficacy should be evaluated based upon reduction in HbA1c (glycosylated haemoglobin) from baseline to end of study, caused by each formulation. HbA1c in normal subjects is typically less than 6.5 and in diabetics, it frequently exceeds 7.0

Obviously, standard formulation is expected to cause slightly greater reduction in HbA1c than the test formulation. However clinicians opined that test formulation can be considered "as good as" standard formulation, if difference in reduction in HbA1c caused by two formulations is not more than 0.4



- A. How will you write the statistical hypothesis for this study [3 marks]
- B. What information will you need to calculate sample size for this study? [3 marks]
- C. Discuss statistical methods that will be used for testing sponsor's objectives in this study. [4 marks]

6. Please answer following multiple choice questions (by indicating on your answer sheet) question number and the option that represents correct answer (e.g. Mention 6a: (i) for question 6a if option numbered (i) is the correct answer in your opinion) [2 Marks Each] (10)

- a. All of the following statements about PROC Freq are true except:
  - (i) Tables statement determines, for which variables, the frequency will be calculated.
  - (ii) Norow, nocol and nopercnt are valid options in Tables statement.
  - (iii) 'Fisher' option in tables statement will calculate Fisher's exact test
  - (iv) This procedure does not produce any visible output
- b. Which of the following is the correct syntax to create a dataset with 100 records numbering from 1 to 100?
  - (i)
 

```
Data ABCD;
A= 1 to 100;
Output;
Run;
```
  - (ii)
 

```
Data ABCD;
A= ranuni(1234)*100;
Output;
Run;
```
  - (iii)
 

```
Data ABCD (drop=i);
Do i = 1 to 100;
A+1;
Output;
End;
Run;
```
  - (iv) Proc SQL;
 

```
Create table abcd as 1to100 from NULL;
Quit;
```

**[TURN OVER**

- c. Identify the missing component of the syntax

Data Exam2;

\_\_\_\_\_ Exam1;

Where Var1 > var2;

Run;

- (i) Merge
- (ii) Set
- (iii) Using
- (iv) Copy

- d. Which of the following is a correct statement about SAS procedures?

- (i) Proc means can provide mean, median and mode, but not quartiles
- (ii) Proc freq can provide frequencies but not column percentages
- (iii) Data step cannot be used to create block randomization
- (iv) Proc univariate can be used to perform tests of normality.

- e. Which one of the following SAS procedure will not display output by default?

- (i) PROC POWER
- (ii) PROC SQL
- (iii) PROC GLM
- (iv) PROC SORT

\*\*\*\*\*



- 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 **both** section should be written in **separate** answer-books.

**Section I**

1. (a) Explain the use of (i) Stars and (ii) Chernoff faces to represent multivariate data. 4

(b) Let  $X = \begin{bmatrix} 8 & 8 & 2 \\ 0 & 2 & 4 \\ 8 & 12 & 10 \end{bmatrix}$  Obtain (i)  $\bar{X}$  (ii)  $S_{n-1}$  (iii) Generalized Sample variance 6

and its interpretation (iv) Total Sample variance.

2. (a) A random sample of 6 observations is selected from a trivariate normal distribution with mean vector. 4

$$\mu = \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} \text{ and } \Sigma = \begin{pmatrix} 4 & 3 & 2 \\ 3 & 1 & 0 \\ 2 & 0 & 9 \end{pmatrix}$$

Obtain distribution of (i)  $\bar{X}$  (ii)  $\sum_{i=1}^6 X_i$ .

- (b) Explain briefly, Transformations to near normality. 3  
(c) Obtain maximum likelihood estimators of the mean vector and variance-covariance matrix using the given sample from a bivariate normal distribution 3

$$X = \begin{bmatrix} 6 & 5 & 4 & 1 \\ 8 & 7 & 8 & 5 \end{bmatrix}$$

Also obtain the sample correlation coefficient matrix.

3. (a) The p.d.f. of a 5-variate normal distribution is as follows :— 4

$$\frac{1}{(2\pi)^{\frac{5}{2}} (42)} \exp \left\{ -\frac{1}{2} \left[ (x_1 - 4)^2 + \frac{(x_2 + 3)^2}{4} + \frac{x_3^2}{7} + \frac{(x_4 - 2)^2}{9} + \frac{(x_5 + 1)^2}{7} \right] \right\}$$

State the mean vector and variance-covariance matrix. Obtain distribution of  $\begin{bmatrix} X_3 \\ X_5 \end{bmatrix}$ .

- (b) A random sample of size 4 from bivariate normal population gave the following :— 6

$$\bar{X} = \begin{bmatrix} 6 \\ 2 \end{bmatrix} \quad S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$$

Compute the  $T^2$  statistic and test the hypothesis  $H_0 : \mu'_0 = [3 \ 3]$ . Obtain the 95% simultaneous confidence intervals for  $\mu_1$  and  $\mu_2$ .  
( $F_{2, 2}(0.05) = 19$ )

4. (a) Explain the use of  $T^2$ -statistic for testing the equality of vector means from two multivariate populations. 5  
(b) Explain the multivariate multiple regression model and state the estimate of the parameters. 5

### Section II

5. (a) Write down orthogonal factor model and explain the terms (i) Communality 4  
(ii) Specific variance. 2  
(b) What is factor rotation ? how it is useful for interpretation ? 2  
(c) Define population principal components. Describe the procedure to obtain principal components from the variance matrix  $\Sigma$ . Obtain variances of the principal components. 4
6. (a) Given the following data compute the sample linear discriminant function and classify the observation  $X_0 = (2 \ 7)'$  using minimum ECM rule with equal priors and equal costs. 5

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

- (b) The distances between pairs of five items are given below :— 5

	1	2	3	4	5
1	0				
2	4	0			
3	6	9	0		
4	1	7	10	0	
5	6	3	5	8	0

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

OR

- (b) Define canonical correlations and canonical variables. Give a test to decide whether it is worthwhile to proceed for canonical correlation analysis. 5