Roll No. | | | | | | | | | | |

ANNA UNIVERSITY (UNIVERSITY DEPARTMENTS)

B.E. (Full Time) –END SEMESTER EXAMINATIONS, NOV / DEC 2023

INDUSTRIAL ENGINEERNG
V Semester
IE5001&Applied Multi-Variate Statistical Analysis
(Regulation 2019)

Time:3hrs Max.Marks: 100

CO1	Predict the values of one or more variables on the basis of observations on the other variables
CO2	Formulate the specific statistical hypotheses, in terms of the parameters of multivariate populations
CO3	Data reduction or structural simplification as simply as possible without sacrificing valuable information and will make interpretation easier
CO4	Sorting and Grouping "similar" objects or variables are created, based upon measured characteristics
CO5	Able to understand appropriate use of methods.

BL – Bloom's Taxonomy Levels
(L1-Remembering, L2-Understanding, L3-Applying, L4-Analysing, L5-Evaluating, L6-Creating)

PART- A(10x2=20Marks)
(Answer all Questions)

Q.No	Questions	Marks	CO	BL
1	List the properties of multivariate normal distribution.	2	1	1
2	Contrast between linear dependent and Independent vectors	2	1	2
3	List the assumption of one-way MANOVA.	2	2	1
4	Contrast between Covariance and correlation	2	2	2
5	What is the use of Scree plot?	2	3	1
6	When will you use Bartlett test of sphericity?	2	3	2
7	List any two types of evaluation criterion used in discriminant analysis and give examples for each.	2	4	3
8	What is the purpose of discriminant analysis?	2	4	2
9	What do you mean by canonical correlation?	2	5	2
10	Differentiate between agglomerative hierarchical and divisive hierarchical clustering algorithm	2	5	2

PART- B(5x 13=65Marks)
(Restrict to a maximum of 2 subdivisions)

Q.No	Questions	Marks	CO	BL
11 (a) (i)	Write the Spectral decomposition for the given symmetric matrix $A=\begin{bmatrix}13 & -4 & 2\\ -4 & 13 & -2\\ 2 & -2 & 10\end{bmatrix}$	13	1	2
	OR			
11 (b) (i)	The following are five measurement on the variable X_1, X_2 and X_3.	8	1	2

Find the arrays of Mean (μ), S_n (Variance . Covariance matrix) and R (Correlation matrix).

X_1	9	2	6	5	8
X_2	12	8	6	4	10
X_3	3	4	0	2	1

(ii) Evaluating T^2, for testing H_0: $\mu=\begin{bmatrix}7\\11\end{bmatrix}$, using the data $X=\begin{bmatrix}2 & 12\\8 & 9\\6 & 9\\8 & 10\end{bmatrix}$ test H_0 at the α =0.05 level. **5** 1 2

13 2 3

12 (a) (i) An article in wear presents data on the fretting wear of mild steel and oil viscosity. Representative data follow with x = oil viscosity and y= wear volume (10^{-4} cubic millimeters) Table 13 a.

Table 13 a

x	1.6	9.4	15.5	20	22	35.5	43.0	40.5	33.0
y	240	181	193	155	172	110	113	75	94

i)Construct a scatter plot of the data.
ii) Fit a simple linear regression model
iii) Test the significance of regression model at a significance level of 0.05.
iv) predict fretting wear when viscosity x=36.0.

OR

12 (b) (i) The annual sales revenue (in crores of rupees) of a product as a function of sales force (number of salesmen) and annual advertising expenditure (in lakhs of rupees) for the past 10 years are summarizes in below Table 12b. Design a regression model to forecast the annual sales revenue of the product **13** 2 3

Table 12 b: Sales Details

Annual sales revenue	20	23	25	27	21	29	22	24	27	35
Sales force	8	13	8	18	23	16	10	12	14	20
Annual advertising Expenditures	28	23	38	16	20	28	23	30	26	32

13 (a) (i) The Marketing Manager of a two-wheeler company designed a questionnaire to study the customers feedback about its two wheeler and in turn he is keen in identifying the factors of his study. He has identified six variables which are as listed below: Fuel efficiency (X_1), life of two wheeler (X_2), Handling convenience(X_3), Quality of original spares(X_4) , Breakdown rate (X_5) and Price(X_6). So the company administered a questionnaire among 10 customers to obtain their opinion on the above variables. The range of score for each variables is assumed to be between 0 and 10. The score "0" means low rating and "10" means high rating. The correlation matrix for the survey data are summarized in Table 13a. Perform factor analysis using centroid method to identify Two factor which can represent the variables of **13** 3 4

the study.

Table 13a: customers feedback Data- Correlation matrix

Respondent	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1	0.742	0.168	0.496	0.484	-0.430
X_2	0.742	1	0.424	0.568	0.474	-0.204
X_3	0.168	0.424	1	0.050	0.238	0.092
X_4	0.496	0.568	0.050	1	0.290	0.196
X_5	0.484	0.474	0.238	0.290	1	0.037
X_6	-0.430	-0.204	0.092	0.196	0.037	1

OR

13 (b) (i) Suppose the random variables Xl, X2 and X3 have the covariance matrix. Calculate the population principal components. **13** 3 3

$$\sum = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

14 (a) (i) The director of a management school wants to do discriminant analysis concerning the effect of two factors, namely, the year spending(in lakhs of rupees) on infrastructures of the school (X_1) and the yearly spending on interface events of the school(X_2) on the grading of the school by an inspection team. He has collected the data for the past 6 years and submitted to the inspection team.Based on the data, The committee has awarded one of the following grades for each year, as shown in Table 14a.Design the Discriminant function- $Y=aX_1 +bX_2$. **13** 4 4

Table – 14a: School Details

Year	Grade	Infrastructure	Interface events
1	Below	3	4
2	Below	4	5
3	Above	10	7
4	Below	5	4
5	Below	6	6
6	Above	11	4

Solve this problem by employing Discriminant Analysis?

OR

14 (b) (i) The Performance standard of employees as a function of their age (X1) and Family size (X2) is classified in to 'Above average' and 'Below Average'. The data on 6 different employees in a company are presented in the following table 14b. **13** 4 3

Table 14b: Observations

Employee	Standard	X1	X2
1	Below	43	3
2	Below	24	4
3	Above	30	6
4	Above	41	3
5	Below	56	5
6	Above	22	4

Design the Discriminant function, Y=aX1 +bX2

15 (a) (i) Six observations on two variables are available, as shown in the following table 15a. **13** 5 3

Table 15a: Observations

Obs	X1	X2
a	-1	-2
b	0	0
c	2	2
d	-2	-2
e	1	-1
f	1	2

i) Plot the observations in a scatter diagram
ii) Apply the Single linkage method and Use a dendrogram to arrive at the number of groups and their membership.
iii) Same as (ii), except apply the furthest neighbor method

OR

15 (b) (i) In a Survey, the number of years of experiences of employees in two different skills are summarized in Table 15b. **13** 5 3

Table 15b: Observations

Employee	Skill – X	Skill-Y
1	2	8
2	8	16
3	3	6
4	6	9
5	8	7
6	10	10

i) Plot the observations in a scatter diagram.
ii) Cluster the employee using agglomerative–Centroid method. Use a dendrogram to arrive at the number of groups and their membership.

PART- C(1x 15=15Marks)

(Q.No.16 is compulsory)

Q.No	Questions	Marks	CO	BL
16.	Observations on two responses are collected for three treatments. The observation vectors $\begin{bmatrix} X1 \\ X2 \end{bmatrix}$ are	**15**	2	5

Treatment I: $\left(\begin{bmatrix} 6 \\ 7 \end{bmatrix}, \begin{bmatrix} 5 \\ 9 \end{bmatrix}, \begin{bmatrix} 8 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ 9 \end{bmatrix}, \begin{bmatrix} 7 \\ 9 \end{bmatrix}\right)$

Treatment 2: $\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}\right)$

Treatment 3: :$\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix}\right)$

i) Break up the observations into mean, treatment, and residual

components, Construct the corresponding arrays for each variable.

ii) Using the information in Part i, construct the one-way MANOVA table and Evaluate Wilks' lambdato test for treatment effects. Set α= .05