


**VIT**

 Vellore Institute of Technology  
Chennai • Vellore • Bangalore • Hyderabad • Pune • Sri Lanka • USA • UK • Australia • Canada • France • Germany • Italy • Japan • Korea • Malaysia • Singapore • Taiwan • Thailand • Vietnam • Philippines • India
**Final Assessment Test - November 2024**

Course: CBS1009 - Computational Statistics

Class NBR(s): 5573 / 5574

Time: Three Hours

Slot: B1

Max. Marks: 100

- KEEPING MOBILE PHONE/ANY ELECTRONIC GADGETS, EVEN IN 'OFF' POSITION IS TREATED AS EXAM MALPRACTICE
- DON'T WRITE ANYTHING ON THE QUESTION PAPER

 Answer ALL Questions

(10 X 10 = 100 Marks)

1. Given the sample variable 1 representing sales 42, 52, 48, and 58 and another variable representing number of books has sample values 4, 5, 4 and 3 then construct the mean vector, covariance vector, correlation matrix and bivariate normal distribution. [10]

2. Let  $X$  be a distributed with  $N_3(\mu, \Sigma)$  where  $\mu^T = [1 \ -1 \ 2]$  and [10]

$\Sigma = \begin{bmatrix} 4 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}$ . Which of the following random variables are independent and explain

- a)  $X_1$  and  $X_2$   
 b)  $X_1$  and  $X_3$   
 c)  $X_2$  and  $X_3$   
 d)  $(X_1, X_3)$  and  $X_2$   
 e)  $X_1$  and  $X_1 + 3X_2 - 2X_3$

$$\begin{bmatrix} 4 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix} \quad \begin{bmatrix} 4 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2.5 \end{bmatrix}$$

3. Fit a Multiple Linear Regression Model for the following dataset [10]

|       |     |     |     |     |     |     |     |     |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|
| Y     | 140 | 155 | 153 | 179 | 192 | 200 | 212 | 215 |
| $X_1$ | 60  | 62  | 67  | 70  | 71  | 72  | 75  | 78  |
| $X_2$ | 22  | 25  | 24  | 20  | 15  | 14  | 14  | 11  |

4. The data represents the Rainfall received during rainy season in Vellore for a period of one week. Find the autocorrelation function for the given dataset lagged by 2 [10]

| Day   | Rainfall (In mm) |
|-------|------------------|
| Day 1 | 27               |
| Day 2 | 28               |
| Day 3 | 31               |
| Day 4 | 32               |
| Day 5 | 37               |
| Day 6 | 30               |
| Day 7 | 24               |

$(X_2, X_3)$   
 $(X_3, X_1) \& X_2$   
 $\begin{matrix} 3 & 1 & 2 \\ 3 & 2 & -1 \\ 2 & -1 & 4 \\ 2 & 0 & 0 \end{matrix}$

5. Consider the following data on one predictor variable  $Z$  and two responses  $Y_1$  and  $Y_2$  [10]

|       |    |    |   |   |   |
|-------|----|----|---|---|---|
| Z     | 0  | 1  | 2 | 3 | 4 |
| $Y_1$ | 1  | 4  | 3 | 8 | 9 |
| $Y_2$ | -1 | -1 | 2 | 3 | 2 |

Determine the least square estimates of the parameters in the multivariate multiple regression linear model.

6. A Doctor wants to examine the effect of three different drugs on two different diseases on 3 different patients and their recovery time is given. The Drugs are classified as : Drug A, Drug B, Drug C. Using MANOVA, Analyze the data and comment it. [10]

77

(A)

| S.No | Drugs | Disease A Recovery Time(In days) | Disease B Recovery Time(In days) |
|------|-------|----------------------------------|----------------------------------|
| 1    | A     | 5                                | 8                                |
| 2    | A     | 9                                | 8                                |
| 3    | A     | 8                                | 10                               |
| 4    | B     | 7                                | 12                               |
| 5    | B     | 8                                | 13                               |
| 6    | B     | 10                               | 9                                |
| 7    | C     | 9                                | 11                               |
| 8    | C     | 9                                | 14                               |
| 9    | C     | 12                               | 16                               |

7. Find the linear discriminant analysis to the following data [10]

$X_1 = (x_1, x_2) = \{(4,1), (2,4), (2,3), (3,6), (4,4)\}$

$X_2 = (x_1, x_2) = \{(4,10), (6,8), (9,5), (8,7), (10,8)\}$

8. The following data represents marks of three students in three different subjects. [10]  
Perform the Principle Component Analysis for the given data

| S.No      | CAT1 Marks (Out of 15) | CAT2 Marks (Out of 15) | Quiz Marks (Out of 10) |
|-----------|------------------------|------------------------|------------------------|
| Student 1 | 10                     | 2                      | 7                      |
| Student 2 | 14                     | 3                      | 8                      |
| Student 3 | 12                     | 3                      | 10                     |

$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

- 9.a) Apply K-mean Clustering for the following data set: [10]  
(10, 12, 14, 15, 16, 23, 25, 26, 32) with 2 clusters.

OR

- 9.b) The distance between pairs of five items are as follows [10]

|   | P | Q | R  | S | T |
|---|---|---|----|---|---|
| P | 0 |   |    |   |   |
| Q | 4 | 0 |    |   |   |
| R | 6 | 9 | 0  |   |   |
| S | 1 | 7 | 10 | 0 |   |
| T | 6 | 3 | 5  | 8 | 0 |

Cluster the five items using the single linkage and draw the dendrograms of the same.

10.a) For the following time series data find the forecast demand for period 6 using

i) Trend Naïve Method

ii) Linear regression model

|          |     |     |     |     |     |
|----------|-----|-----|-----|-----|-----|
| Week t   | 1   | 2   | 3   | 4   | 5   |
| Demand Y | 150 | 157 | 162 | 166 | 177 |

OR

10.b) The following data represents the marks of 10 students in FAT examination of a Statistics subject. [10]

| Sl. No | FAT Marks (Statistics ) |
|--------|-------------------------|
| 1      | 78                      |
| 2      | 67                      |
| 3      | 87                      |
| 4      | 69                      |
| 5      | 89                      |
| 6      | 95                      |
| 7      | 91                      |
| 8      | 90                      |
| 9      | 56                      |
| 10     | 83                      |

Calculate the 3 moving average and 5 moving average of the above data.

⇒⇒⇒BG/K/TX⇒⇒⇒

$L_{Df_2} = (S_w)^{-1} \times \begin{pmatrix} u_1 & u_2 \end{pmatrix}$