


VIT

Vellore Institute of Technology

Final Assessment Test – November 2025

Course: CBS1009 - Computational Statistics

Class NBR(s): 4403 / 4405

Time: Three Hours

Slot: A1

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

General Instructions: Statistical Tables are permitted.

COs	CO Statements
CO1	Analyse and interpret statistical data using multivariate normal distributions.
CO2	Learn the approaches to point estimation of parameters.
CO3	Understand the concept of multivariate regression, by using multivariate analysis and interpreting experimental data.
CO4	Understand the concept of statistical analysis.
CO5	Learn about the data aggregation, group operations and time series.

BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)
Answer ALL Questions
(10 X 10 = 100 Marks)

- Two sensors - LIDAR (X) and RADAR (Y) - are used to estimate a vehicle's distance from an obstacle. The measurements from both sensors are assumed to follow normal distribution with the following parameters:
Mean: $\mu_x = 10$ m, and $\mu_y = 9.8$ m ;
Standard deviation: $\sigma_x = 0.5$ and $\sigma_y = 0.8$;
Correlation coefficient: $\rho = 0.9$.
i) Find the squared distance between the two sensors?
ii) Identify and sketch the constant- density contour that encloses 90% of the total probability of the joint distribution?
CO1 BL2
- Inside a chemical reactor the triplet of measurements - temperature X_1 (°C), pressure X_2 (kPa), and humidity X_3 (%) – follows a trivariate normal distribution with mean vector and covariance matrix as:

$$\mu = \begin{pmatrix} 25 \\ 100 \\ 60 \end{pmatrix}, \Sigma = \begin{bmatrix} 4 & 2 & 0 \\ 2 & 9 & 0 \\ 0 & 0 & 16 \end{bmatrix}$$

i) Find the conditional probability $\Pr[X_2 > 104 / X_1 = 32^\circ\text{C}]$.
ii) Are the two random variables $U = X_1 - 2X_2$ and $V = 2X_2 + X_3 - X_1$ independent?
CO1 BL2
- The following data were recorded by a researcher from six experimental plots to investigate how rainfall (X_1 in mm) and fertilizer application (X_2 , in kg/ha) influences the wheat crop yield (Y , in kg/ha) .
CO2 BL3

Plot	X_1 (Rainfall, mm)	X_2 (Fertilizer, kg/ha)	Y (Yield, kg/ha)
1	85	40	120
2	90	50	135
3	78	45	115
4	92	55	140
5	88	48	130
6	80	42	118

- i) Using the given data, estimate the parameters of the multiple linear regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ by the method of least squares.
- ii) Using the fitted model, predict the wheat yield when the rainfall is 89 mm and fertilizer is 50 kg/ha.

CO2 BL3

4. A company tests 4 battery samples and record three measurements for each X_1 (charging time in hours), X_2 (Operating time in hours) and X_3 (recharge cycle in hundreds). The data assumed to follow trivariate normal distribution and the sample data set is given below (each column represents on trial or one battery)

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

- i) Compute the maximum likelihood estimate (MLE) of the mean vector and the covariance matrix.
- ii) Based on the covariance estimates, comment on whether batteries with longer charging time lend to have longer operating time?

5. A Gym evaluates three exercise programs (Cardio, Strength, Mixed) on two outcomes: VO_2 Max Increase (ml/kg/min) and Muscle Gain (kg) after 8 weeks. Using MANOVA, analyse the data to find which program shows balanced improvement across both outcome.

CO3 BL3

Program	VO_2 Max	Muscle Gain
Cardio	5, 6, 4, 5	1, 1.2, 1.1, 1.3
Strength	2, 3, 2.5, 3	3, 3.2, 3.1, 3.3
Mixed	4, 4.5, 5, 4.8	2, 2.1, 2.2, 2.0

6. A political analyst examines how the amount of media exposure (X_1 , hours/week) affects public support for two political parties in a region. The analyst measures the support scores (out of 100) for Party A (Y_1) and Party B (Y_2). Using media exposure as the predictor variable and the party support scores as the response variables, estimate the respective regression equations. Determine whether media exposure has a stronger influence on support for Party A or Party B?

CO3 BL33

X_1	10	12	15	11	14
Y_1	60	65	70	62	68
Y_2	45	48	50	46	49

7. Doctors want to classify patients as Healthy or Diabetic based on Fasting Blood Sugar (FBS, mg/dL) and BMI (kg/m²)

CO4 BL3

Id	Fasting Blood Sugar	BMI	Status
1	85	22	Healthy
2	90	24	Healthy
3	88	23	Healthy
4	140	30	Diabetic
5	150	32	Diabetic
6	155	31	Diabetic

Using discriminant function analysis find out the projection direction that maintains separability between the two classes: Healthy and Diabetic. Classify a new patient with Fasting Blood Sugar = 120, BMI = 28.

8. A smartwatch tracks three health related metrics from users are observed and found the correlation matrix corresponding to the random variables Heart Rate (bpm), Sleep Hours and Stress level (1- 10) is given in CO4 BL3

$$R = \begin{bmatrix} 1 & -0.84 & 0.79 \\ -0.84 & 1 & -0.63 \\ 0.79 & -0.63 & 1 \end{bmatrix}$$

- i) Find the principal components Y_i , $i = 1, 2$, and 3 and loadings.
 ii) Calculate the proportion of total variance explained by the first two principal components to understand the main factors affecting wellness.

- 9.a) A library wants to group documents (D1, D2, D3, D4 and D5) based on their content similarity, computed using text distance matrix. CO4 BL3

$$\text{Distance matrix } D = \begin{bmatrix} 0 & 5 & 9 & 4 & 8 \\ 5 & 0 & 7 & 3 & 6 \\ 9 & 7 & 0 & 8 & 3 \\ 4 & 3 & 8 & 0 & 7 \\ 8 & 6 & 3 & 7 & 0 \end{bmatrix}$$

- i) single linkage hierarchical procedure,
 ii) Average linkage hierarchical procedure,
 iii) Centroid linkage hierarchical procedure. Draw the dendrograms and compare the results in (i), (ii), and (iii).

OR

- 9.b) Consider $p = 3$ and $m = 1$, and suppose the random variables X_1 , X_2 and X_3 have the positive definite covariance matrix defined as follows: CO4 BL3

$$\Sigma = \begin{bmatrix} 1.0 & 0.63 & 0.45 \\ 0.63 & 1.0 & 0.35 \\ 0.45 & 0.35 & 1.0 \end{bmatrix}$$

Check if the proper solution for the factor model exists.

- 10.a) Determine the stationarity of the following two-dimensional vector model CO5 BL3
 $(I - \Phi B)Z_t = a_t$, where $\Phi = \begin{pmatrix} 0.8 & 0.6 \\ -0.2 & 0.7 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 2 \end{pmatrix}$.
 Find the first three matrices $\Gamma(0)$, $\Gamma(1)$, and $\Gamma(2)$.

OR

- 10.b) A retail chain records product sales: CO5 BL3

Region	Product	Sales
East	Pen	120
East	Book	200
West	Pen	150
West	Book	300
North	Pen	100
North	Book	250

- i) Find the average sales Per product and per region.
 ii) Transform Sales to z scores within each region.
 iii) Find for each product's % contribution to region sales total.

⇔⇔⇔ Q/K/TY ⇔⇔⇔