Register Number: ____________________
Date: ____________________

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BENGALURU-27
BCA(DATA ANALYTICS) —III SEMESTER
SEMESTER EXAMINATION: OCTOBER 2021
(Examination conducted in January —March 2022)
BCADA 3121: MULTIVARIATE STATISTICS

TIME: 2.5 Hrs **MAXIMUM MARKS: 70**

This paper has 3 printed pages and 3 parts.

PART 1

Answer all questions. More than one options may be correct. $(\mathbf{1} \times \mathbf{20} = \mathbf{20})$

1. Which of the following are linearly independent sets of vectors?
   A. {(1,1),(0,1)} B. {(2,3,4),(4,6,8)} C. {(0,0,1),(0,1,0),(0,0,0)} D. {(1,1,0),(1,0,1),(0,1,1)} (1)

2. Inverse of the matrix $A = \begin{bmatrix} 3 & 6 \\ 1 & 5 \end{bmatrix}$ is:
   A. $A^{-1} = \frac{1}{9}\begin{bmatrix} 5 & -6 \\ -1 & 3 \end{bmatrix}$ B. $A^{-1} = \frac{-1}{9}\begin{bmatrix} -5 & 6 \\ 1 & -3 \end{bmatrix}$ C. $A^{-1} = \begin{bmatrix} \frac{3}{9} & \frac{6}{9} \\ \frac{1}{9} & \frac{5}{9} \end{bmatrix}$ D. $A^{-1} = \begin{bmatrix} 3 & 6 \\ 1 & 5 \end{bmatrix}$

3. Let $\mathbf{X} \sim N_7(\mu, \mathbf{\Sigma})$, with $\hat{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 & \mu_6 & \mu_7 \end{bmatrix}$ and $\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma 12 & \sigma_{13} & \sigma_{14} & \sigma_{15} & \sigma_{16} & \sigma_{17} \\ \sigma_{21} & \sigma 22 & \sigma_{23} & \sigma_{24} & \sigma_{25} & \sigma_{26} & \sigma_{27} \\ \sigma_{31} & \sigma 32 & \sigma_{33} & \sigma_{34} & \sigma_{35} & \sigma_{36} & \sigma_{37} \\ \sigma_{41} & \sigma 42 & \sigma_{43} & \sigma_{44} & \sigma_{45} & \sigma_{46} & \sigma_{47} \\ \sigma_{51} & \sigma 52 & \sigma_{53} & \sigma_{54} & \sigma_{55} & \sigma_{56} & \sigma_{57} \\ \sigma_{61} & \sigma 62 & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_{66} & \sigma_{67} \\ \sigma_{71} & \sigma 72 & \sigma_{73} & \sigma_{74} & \sigma_{75} & \sigma_{76} & \sigma_{77} \end{bmatrix}$.

   Then, for $\mathbf{X_1} = \begin{bmatrix} X_2 \\ X_5 \\ X_6 \end{bmatrix}$ a subvector of $\mathbf{X}$, the mean vector is:
   A. $\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$ B. $\begin{bmatrix} \mu_5 \\ \mu_6 \\ \mu_2 \end{bmatrix}$ C. $\begin{bmatrix} \mu_2 \\ \mu_5 \\ \mu_6 \end{bmatrix}$ D. $\begin{bmatrix} \mu_6 \\ \mu_2 \\ \mu_5 \end{bmatrix}$ (1)

4. Considedr the setting of question 3. Then, $\mathbf{\Sigma}$ for $\mathbf{X_1}$ is given by:
   A. $\begin{bmatrix} \sigma_{22} & \sigma_{25} & \sigma_{26} \\ \sigma_{25} & \sigma_{55} & \sigma_{56} \\ \sigma_{26} & \sigma_{56} & \sigma_{66} \end{bmatrix}$ B. $\begin{bmatrix} \sigma_{22} & \sigma_{25} & \sigma_{26} \\ \sigma_{52} & \sigma_{55} & \sigma_{56} \\ \sigma_{62} & \sigma_{65} & \sigma_{66} \end{bmatrix}$ C. $\begin{bmatrix} \sigma_{22} & \sigma_{25} & \sigma_{26} \\ \sigma_{15} & \sigma_{55} & \sigma_{56} \\ \sigma_{16} & \sigma_{65} & \sigma_{66} \end{bmatrix}$ D. $\begin{bmatrix} \sigma_{11} & \sigma_{25} & \sigma_{26} \\ \sigma_{52} & \sigma_{22} & \sigma_{56} \\ \sigma_{62} & \sigma_{65} & \sigma_{33} \end{bmatrix}$ (1)

5. In simple regression, the least square estimators are unbiased estimators of the model parameters.
   A. True B. False (1)

6. In a regression model, we assume the expectation of the error term to be:
   A. 0 B. A constant C. mean of the error terms. D. sample mean of the predicted variable. (1)

7. We can use the likelihood ratio test to asses the effect of particular variables on response variables.
   A. True B. False (1)

8. In a trivariate case, the partial correlation coefficient between two variable (say X_1, X_3) is:
   A. Correlation between X_1 and X_2. B. Effect of X_1 on X_2.
   C. Correlation between X_1 and X_2 after eliminating the linear effect of third variable on X_1 and X_2. (1)

9. In factor analysis, how do we choose the number of factors?
   A. Scree Plot B. Eigen Values C. Trial Error D. It's a random choice. (1)

10. One reason factor roations is done to increase interpretability.
    A. True B. False (1)

11. Principal component analysis is done to
    A. Reduce Dimensions B. Find common factors between the variables.
    C. Find out which variable has the most affact on the model. (1)

12. Factor Analysis is done to
    A. Reduce Dimensions B. FInd common factors between the variables.
    C. Find out which variable has the most affact on the model. (1)

13. In a two class classification problem, we can minimize the expected cost to define the classification regions.
    A. True B. False (1)

14. Let $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ be the probability density functions associated with the random variable $\mathbf{X}$ for the populations π_1 and π_2, respectively. Let Ω be the sample space. R_1 and R_2 be the set of $\mathbf{x}$ values for which we classify objects as π_1 and π_2 respectively. Then the conditional probability of classifying an object as π_2 when it is from π_1 is:
    A. $\int_{R_2} f_2(\mathbf{x})\,dx$ B. $\int_{R_2} f_1(\mathbf{x})\,dx$ C. $\int_{R_1} f_1(\mathbf{x})\,dx$ D. $\int_{R_1} f_2(\mathbf{x})\,dx$ (1)

15. Equivalent linear combinations lead to the same value of the (univariate) standard distance between two points.
    A. True B. False (1)

16. K means is an example of heirarchial clustering.
    A. True B. False (1)

17. Minkowski metric is a distance measure given by:
    A. $[\sum_{i=1}^{p} |x_i - y_i|^m]^{\frac{1}{m}}$ B. $[\sum_{i=1}^{p} |x_i + y_i|^m]^{\frac{1}{m}}$ C. $[\sum_{i=1}^{p} (x_i - y_i)^m]^{\frac{1}{m}}$ D. $[\sum_{i=1}^{p} (x_i + y_i)^m]^{\frac{1}{m}}$ (1)

18. For what value of m does the Minkowski metric give the eucledian distance?
    A. 0 B. 1 C. 2 D. 3

19. Clustering can be heirarchial or non heirarchial.
    A. True B. False

20. Standard can be used to measure the overlap of two normal populations wirh shared variances and different means in discriminant analysis.
    A. True B. False

PART B

Answer ANY SIX questions. **($6 \times 5 = 30$)**

1. Explain how to calculate the eigenvectors and eigenvalues of a matrix. Calculate the eigenvalues and eigenvectors for $\begin{bmatrix} 3 & 4 \\ 1 & 7 \end{bmatrix}$ (5)

2. Explain with an example how we can find the distribution of linear combinations of Multivariate normal distributed variables. (5)

3. For the following multuvariate data, fit a linear regression model. Here, Y_1, Y_2 are the response variables. (5)

Z	15	9	3	25	7	13
Y_1	10	5	7	19	11	18
Y_2	2	3	3	6	7	9

4. Describe the model for factor analysis. (5)

5. How can we determine the number of factors to be used a factor analysis model? (5)

6. Explain finding classification regions using ECM for two class classification. (5)

7. Breifly explain one heirarchial and one non heirarchial clustering method. (5)

8. Briefly explain when we used the likelihood ratio test, and for what. (5)

PART C

Answer ANY TWO questions. $(\mathbf{2} \times \mathbf{10} = \mathbf{20})$

1. Explain single and multiple regression models. What are the assumptions? When do we use these? (10)

2. Explain in detail the procedure for factor analysis. (10)

3. Suppose that $\mathbf{X}$ follows bivariate distribution with $E_1[\mathbf{X}] = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in group 1 and $E_2[\mathbf{X}] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ in group 2, and common covariance matrix
$psi = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $-1 < \rho < 1$.
   (a) Compute the vector of discriminant function coefficients. How does the discriminant function depend on ρ?.
   (b) Compute the bivariate standard distance as a function of ρ. What are the minimum and maximum of this function? For which values of ρ are they attained?

(10)