Manipal Academy of Higher Education

Second Semester ME (BIG DATA AND DATA ANALYTICS)

BDA –616.2: Applied Multivariate Analysis

Scheme for Evaluation

1.

(a) Discuss the importance of multivariate analysis in big data analytics. Define a sample covariance matrix.

(3+2 = 5 marks)

- **Any 3 importance of multivariate analysis – 3 Marks**
- **Definition of sample covariance matrix – 2 Marks**

(b) Let $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$ be a multivariate normal random vector with mean vector $\mu = \begin{bmatrix} 20 \\ 13 \\ 23 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 67 & 29 & 76 \\ 29 & 85 & 90 \\ 76 & 90 & 50 \end{bmatrix}$. Find mean and variance of $Z = 10Y_1 + 2Y_2 - 4Y_3$.

(2+3=5 Marks)

- **Correct answer for mean with method – 2 Marks**
- **Correct answer for variance with method – 3 Marks**

2. Discuss the data structure, model, hypothesis and between and within variance structure in multivariate one-way analysis of variance (MANOVA). Also mention the names of different test statistics used in MANOVA

(2+2+1+3+2=10 Marks)

- **Data structure of MANOVA – 2 Marks**
- **Model – 2 Marks**
- **Hypothesis – 1 Mark**
- **Between and within variance structure – 3 Marks**
- **Names of 4 test statistics used in MANOVA – 2 marks**

3. Describe principal component analysis (PCA) for ‘p’ number of variables? Mention the different methods used for estimating factor loadings and communalities in factor analysis

(6+4 = 10 marks)

- **Introduction – 2 Marks**
- **Describing the method of deriving the PCA – 4 Marks**
- **Names of 4 different methods used in factor analysis – 4 Marks**

4. Write the similarities and dissimilarities between principal component analysis and exploratory factor analysis. Write short notes on methods on deciding how many principal components to be retained

(6+4=10 Marks)

- **Any 5 similarities or dissimilarities – 6 marks**
- **4 different methods on deciding how many principal components to be retained – 4 Marks**

5. What do you mean by rotation in factor analysis and discuss its uses? Explain different type of rotation used in factor analysis

(4+6 = 10 marks)

- **Definition and use of rotation in factor analysis – 4 marks**
- **Two types of rotation – 3+3= 6 Marks**

6. Describe the discriminant analysis for two group cases

(10 arks)

- **Need and introduction to Discriminant analysis – 2 Marks**
- **Assumptions – 3 Marks**
- **Derivation of the Discriminant function – 5 Marks**

7. Describe the classification analysis for two group cases

(10 Marks)

- **Need and introduction Classification analysis– 2 Marks**
- **Assumptions – 3 Marks**
- **Derivation of the function – 5 Marks**

8. Explain the divisive and agglomerative method of identifying clusters in cluster analysis

(5+5=10 Marks)

- **Divisive method – 5 Marks**
- **Agglomerative method – 5 Marks**

9. What do you mean by hierarchical clustering? Explain the average linkage and centroid approaches to measure the distance between clusters.

(2+4+4=10 Marks)

- **Hierarchical clustering - 2 Marks**
- **Average linkage Method – 4 Marks**
- **Centroid Approach – 4 Marks**

10. Differentiate between cluster analysis and discriminant analysis

(10 Marks)

- **At least 3 point in detail – 10 Marks**

****