

Enrollment No.....



Faculty of Engineering
End Sem (Odd) Examination Dec-2022
EN3BS09 Computational Statistics

Programme: B.Tech.

Branch/Specialisation: CSBS

Duration: 3 Hrs.**Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d.

- Q.1 i. Which average is affected most by extreme observation? 1
 (a) Median (b) Arithmetic mean
 (c) Harmonic mean (d) Geometric mean
- ii. The type of data on which analysis is based on more than two variables. 1
 (a) Univariate data (b) Bivariate data
 (c) Multivariate data (d) None of these
- iii. In Probability density of normal distribution of x. What is x? 1
 (a) Random variable (b) Mean value of variable
 (c) Variance of the variable (d) Unknown variable
- iv. The random variable X which has mean as μ and variance as σ^2 can be written as- 1
 (a) $X \sim N(\mu, \sigma^2)$ (b) $X \sim N(\sigma^2, \mu')$
 (c) $X \sim N(\sigma, \mu')$ (d) $X \sim N(\mu, \sigma)$
- v. Eigenvalue and eigenvector are used to- 1
 (a) Determine the number of factors
 (b) Estimation of the loading factor
 (c) Estimation of specific variance
 (d) All of these
- vi. The number of eigenvalues of n x n matrix is- 1
 (a) 1 (b) 2 (c) 3 (d) n
- vii. A factor loading of 0.80 means- 1
 (a) There is no relationship between that variable and the factor
 (b) The item correlates well with the factor, though not perfectly
 (c) The item is poorly related to the factor
 (d) The variable is moderately related with the factor

P.T.O.

[2]

- viii. A factor loading is- **1**
 (a) A correlation coefficient between a variable and a factor (cluster of variables)
 (b) Empirically based hypothetical variable consisting of items which are strongly associated with each other.
 (c) The correlation between a binomial variable and a variable which has a continuous distribution of scores.
 (d) The correlation of a variable with a whole score
- ix. Which of the following is finally produced by hierarchical clustering? **1**
 (a) Final estimate of cluster centroids
 (b) Tree showing how close things are to each other
 (c) Assignment of each point to clusters
 (d) All of these
- x. Point out the correct statement- **1**
 (a) The choice of an appropriate metric will influence the shape of the clusters
 (b) Hierarchical clustering is also called HCA
 (c) In general, the merges and splits are determined in a greedy manner
 (d) All of these
- Q.2 i. Explain the difference between variance and standard deviation. **2**
 ii. What is the empirical formula with related to normal distribution data? **3**
 Explain with an example.
 iii. What is multivariate normal distribution? Explain with an example. **5**
- OR iv. How estimation of parameters is accomplished? Explain with an example. **5**
- Q.3 Attempt any two:
 i. What is the difference between the multiple regression model and the multivariate regression model? **5**
 ii. Explain different types of multivariate regression model. **5**
 iii. What are the basic assumptions required to validate the regression model? **5**
- Q.4 i. Explain two group discriminant analysis. **4**

[3]

- ii. Explain principal component analysis with an example. Also, explain the importance of eigenvalues for determining the dimension or features. **6**
- OR iii. Explain fisher linear discriminant analysis. State the algorithm with an example. **6**
- Q.5 i. Explain factor analysis with an example. **4**
 ii. How common factors extracted? Elaborate with some examples. **6**
- OR iii. What is factor transformation? Why it is needed? **6**
- Q.6 Attempt any two:
 i. How correlation and distance measurement help for cluster analysis? **5**
 ii. Explain K-Means clustering with an example. **5**
 iii. Explain hierarchical clustering with an example. **5**

Marking Scheme
EN3BS09 Computational Statistics

Q.1	i)	Which average is affected most by extreme observation? a) Median b) Arithmetic mean (ANSWER) c) Harmonic mean d) Geometric mean	1
	ii)	The type of data on which analysis is based on more than two variables. a) Univariate Data b) Bivariate Data c) Multivariate Data (ANSWER) d) None of these.	1
	iii)	In Probability density of normal distribution of x . What is x ? a) Random Variable (ANSWER) b) Mean value of Variable c) Variance of the variable d) Unknown variable	1
	iv)	The random variable X which has mean as μ and variance as σ^2 can be written as: a) $X \sim N(\mu, \sigma^2)$ (ANSWER) b) $X \sim N(\sigma^2, \mu')$ c) $X \sim N(\sigma, \mu')$ d) $X \sim N(\mu, \sigma)$	1
	v)	Eigenvalue and eigenvector are used to a) Determine the number of factors b) Estimation of the loading factor c) Estimation of specific variance d) All of the above (ANSWER)	1
	vi)	The number of eigenvalues of n x n matrix is: a) 1 b) 2 c) 3 d) n (ANSWER)	1
	vii)	A factor loading is: a) A correlation coefficient between a variable and a factor (cluster of variables). (ANSWER) b) Empirically based hypothetical variable consisting of items which are strongly associated with each other. c) The correlation between a binomial variable and a variable which has a continuous distribution of scores.	1

		d) The correlation of a variable with a whole score.	
	vii)	A factor loading of 0.80 means: a) There is no relationship between that variable and the factor. b) The item correlates well with the factor, though not perfectly. (ANSWER) c) The item is poorly related to the factor. d) The variable is moderately related with the factor.	1
	ix)	Which of the following is finally produced by Hierarchical Clustering? a) Final estimate of cluster centroids b) Tree showing how close things are to each other (ANSWER) c) Assignment of each point to clusters d) All of the above mentioned	1
	x)	Point out the correct statement. a) The choice of an appropriate metric will influence the shape of the clusters b) Hierarchical clustering is also called HCA c) In general, the merges and splits are determined in a greedy manner d) All of the mentioned (ANSWER)	1
Q.2	i.	Explain the difference between Variance and Standard Deviation.	1+1
	ii.	What is the Empirical formula with related to Normal distribution data, explain with an example	2+1
	iii.	What is Multivariate Normal Distribution, explain with an example.	3+2
OR	iv.	How estimation of parameters is accomplished, explain with an example	3+2
Q.3	i.	What is the difference between the multiple regression model and the multivariate Regression model? (2.5 +2.5)	5
	ii.	Explain different types of Multivariate regression model	5
OR	iii.	What are the basic assumptions required to validate the regression model? (5 POINTS-5 MARKS)	5
Q.4	i.	Explain two group discriminant analysis. Group Name and explanation Example	3+1 3 marks 1 mark
	ii.	Explain Principal Component Analysis, with an example. Also, explain the importance of Eigenvalues for determining the dimension or features.	4+2
OR	iii.	Explain Fisher Linear Discriminant Analysis, State the algorithm with an example.	3+3

Q.5	i.	Explain Factor analysis, explain with an example?	3+1
	ii.	How common factors extracted, elaborate with some examples.	4+2
OR	iii.	What is factor transformation and why it is needed.	4+2
Q.6		Attempt any two:	
	i.	How Correlation and Distance measurement help for Cluster Analysis	5
	ii.	Explain K-Means Clustering with an example.	5
	iii.	Explain hierarchical Clustering with an example.	5
