Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering
End Sem Examination Dec-2023
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. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. Standard Normal Distribution has which of the following properties: **1**

(a) Mean=Variance=1
(b) Standard deviation=Variance=0
(c) Mean=0, Variance=1
(d) Mean=Standard deviation

ii. The shape of Normal distribution data curve is_____: **1**

(a) Bell shaped (b) Triangular (c) Rectangular (d) Circular

iii. Which of the following methods is used to select the best subset of predictor variables in multiple linear regression? **1**

(a) Forward selection (b) Backward elimination
(c) Stepwise regression (d) All of these

iv. Which analysis is portrayed by the equation: $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3... + bnXn$? **1**

(a) Simple regression (b) Multiple regression
(c) Chi-square (d) None of these

v. What is the condition of multicollinearity? **1**

(a) Two or more independent variables are highly correlated
(b) Two or more dependent variables are highly correlated
(c) The independent variables are not correlated with the dependent variable
(d) None of these

vi. Generally, Principal Component Analysis can be applied on: **1**

(a) Labelled dataset (b) Unlabelled dataset
(c) Both (a) & (b) (d) Datasets not required

vii. A multivariate statistical technique for studying interrelationships among variables, usually for discovering underlying constructs or data reduction is known as- **1**

(a) Multiple regression (b) Factor analysis
(c) Discriminant analysis (d) None of these

viii. In exploratory factor analysis, how much variance would a good model be likely to explain? **1**

(a) 0 to 25% (b) 25% to 50% (c) 50% to 75% (d) 75% to 100%

ix. Which of the following is required by K-means clustering? **1**

(a) Defined distance metric (b) Number of clusters
(c) Initial guess as to cluster centroids (d) All of these

x. Which of the following clustering requires merging approach? **1**

(a) Partitional (b) Hierarchical
(c) Naive Bayes (d) None of these

Q.2 i. Explain the term "Estimation of Parameters". **2**

ii. Mention the conditions under which Binomial distribution tends to Normal distribution. **3**

iii. The marks of 1000 students in an exam follows Normal distribution with Mean 70 & Std. Deviation 5. Find number of students whose marks will be- **5**

(a) Less than 65
(b) More than 75
(c) Between 65 & 75

Table: Standard Normal Cumulative Probabilities (Z-score)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177

OR iv. How do you determine the best fit line for a linear regression model? **5**

Q.3 i. What is the difference between the simple linear regression model and the multiple linear regression model? **4**

ii. Explain multivariate analysis of variance and covariance. **6**

OR iii. What are the basic assumptions required to validate the regression model? **6**

Q.4 i. Explain discriminant analysis with an example. **3**

ii. Consider the two-dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8). Compute the principal component using PCA Algorithm. **7**

OR iii. What is fisher linear discriminant analysis? Explain computing process to find eigen vectors using linear discriminant analysis. **7**

Q.5 i. What is difference between factor analysis and normality test? **4**

ii. What do you mean by factor loading? Explain with an example. **6**

OR iii. How common factors extracted? Elaborate with some examples. **6**

Q.6 Attempt any two:

i. What is cluster analysis? Explain its types. **5**

ii. Differentiate between hierarchical and overlapping clustering. **5**

iii. Explain K-Means clustering with an example. **5**

******

Marking Scheme

Computational Statistics- EN3BS09 (T)

Q.1	i)	c) Mean=0, Variance=1 (Answer)		**1**
	ii)	a) Bell shaped (Answer)		**1**
	iii)	d) All of the Above (Answer)		**1**
	iv)	b) Multiple regression (Answer)		**1**
	v)	a) Two or more independent variables are highly correlated (Answer)		**1**
	vi)	b) Unlabelled dataset (Answer)		**1**
	vii)	b) Factor analysis (Answer)		**1**
	viii)	c) 50% to 75% (Answer)		**1**
	ix)	d) all of the mentioned (Answer)		**1**
	x)	b) Hierarchical (Answer)		**1**
Q.2	i.	"Estimation of Parameters".	(As per explanation)	**2**
	ii.	Mention the conditions	(1 Mark*3)	**3**
	iii.	i) Less than 65 (Ans=159)	2 Marks	**5**
		ii) More than 75 (Ans=159)	2 Marks	
		iii) Between 65 & 75 (Ans=682)	1 Mark	
OR	iv.	Definition	2 Marks	**5**
		Example	3 Marks	
Q.3	i.	The Simple Linear Regression model	2 Marks	**4**
		The Multiple Linear Regression model	2 Marks	
	ii.	Analysis of variance	3 Marks	**6**
		Covariance.	3 Marks	
OR	iii.	Basic assumptions…….. model	(1 Mark*6)	**6**
Q.4	i.	Definition	2 Marks	**3**
		Example	1 Mark	
	ii.	Eigen Value	4 Marks	**7**
		Eigen Vector	3 Marks	
OR	iii.	Definition	3 Marks	**7**
		Steps	4 Marks	
Q.5	i.	Difference between factor analysis	2 Marks	**4**
		Normality test	2 Marks	
	ii.	Factor Loading	3 Marks	**6**
		Explain with an example	3 Marks	
OR	iii.	Factors extracted with some examples	3 Marks	**6**
		Elaborate with some examples.	3 Marks	
Q.6		Attempt any two:		
	i.	Cluster Analysis (Definition)	2 Marks	**5**
		Explain its types.	3 Marks	
	ii.	Differentiate between Hierarchical	2.5 Marks	**5**
		Overlapping clustering.	2.5 Marks	
	iii.	Definition	2 Marks	**5**
		Examples	3 Marks	

******