Subject Code:- BCSBS0301 Roll. No: (An Autonomous Institute Affiliated to AKTU, Lucknow) **B.Tech** SEM: III - THEORY EXAMINATION (2024 - 2025) Subject: Computational Statistics Time: 3 Hours General Instructions: IMP: Verify that you have received the question paper with the correct course, code, branch etc. 1. This Question paper comprises of three Sections -A, B, & C. It consists of Multiple Choice *Questions (MCQ's) & Subjective type questions.* 2. Maximum marks for each question are indicated on right -hand side of each question. 3. Illustrate your answers with neat sketches wherever necessary. 4. Assume suitable data if necessary. 5. Preferably, write the answers in sequential order. 6. No sheet should be left blank. Any written material after a blank sheet will not be evaluated/checked. **SECTION-A** 1. Attempt all parts:-1-a. All of the following are examples of dependence methods of analysis EXCEPT? [CO1]

- multiple regression analysis (a)
- multiple discriminant analysis (b)
- multivariate analysis of variance (c)
- (d) cluster analysis

1-b. The off-diagonal elements in the variance-covariance matrix contain: ? [CO1]

- (a) **Regression parameters**
- Means (b)
- (c) Covariance estimates
- Variance estimates (d)
- 1-c. The goal of discriminant analysis is ? [CO2]
 - to develop a model to predict new dependent values (a)
 - the develop a rule for predicting to what group a new observation is most likely to (b) belong

(c) to develop a rule for predicting how independent variable values predict dependent values.

(d) none of these

Printed Page:-04

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

Max. Marks: 100

20

1

1

1

1-d. The coefficient of determination is ? [CO2]

- (a) r
- r^2 (b)
- (c) Ζ
- None of these (d)

1-e. What is the main advantage of a histogram compared to a bar chart? [CO3] 1

1

1

1

1

1

1

- Histograms are better for showing the relationships between variables. (a)
- Bar charts are better for showing the distribution of a single variable. (b)
- (c) Histograms can be used to display both categorical and numerical data.
- (d) Histograms provide a better visual representation of data distribution.
- 1-f. PCA helps in reducing the dimensionality of data by:? [CO3]
 - (a) Discarding important features
 - Keeping all the original features (b)
 - Creating new features (c)
 - Selecting a subset of the original features (d)
- 1-g. Which statistical method is commonly used to calculate factor scores in factor c-2024 analysis? [CO4]
 - Principal Component Analysis (PCA) (a)
 - (b) **T-Test**
 - (c) Multiple Regression
 - None of them (d)
- In factor analysis, what is communality? [CO4] 1-h.
 - The degree to which a variable is related to other variables (a)
 - The shared variance between factors (b)
 - The total variance explained by a factor (c)
 - The number of observations in the dataset (d)
- 1-i. Which clustering method creates a hierarchical representation of data by successively merging or splitting clusters, allowing for a tree-like structure of clusters? [CO5]
 - **K-Means Clustering** (a)
 - Agglomerative Clustering (b)
 - DBSCAN (c)
 - None of them (d)

How does the K-Means algorithm initialize cluster centroids? [CO5] 1-j.

- andomly selects initial centroids (a)
- Uses the farthest data point as the initial centroid (b)
- Finding by standard deviation of data (c)
- None of them (d)

- 2. Attempt all parts:-
- 2 2.a. Define multivariate analysis? [CO1] 2 2.b. What is the use of ANOVA and how can it be useful in analysis of Data? [CO2] 2.c. What do you understand by poor quality of data? [CO3] 2 What are the two primary purposes of factor analysis? [CO4] 2.d. 2 2 2.e. Why is the need to calculate the Euclidean distance in clustering? [CO5] **SECTION-B** 30 3. Answer any five of the following:-What is parameter estimation. Define efficiency and sufficiency of an 3-a. 6 estimator? [CO1] 3-b. If X distributed as $N_4(\mu, \Sigma)$, where? [CO1] 6
- - 3 -1**#** 5 $\begin{bmatrix} 11 & -8 & 3 & 9 \\ -8 & 9 & -3 & 6 \\ 3 & -3 & 2 & 3 \\ 2 & 6 & 3 & 9 \end{bmatrix}$

i) Find the distribution of Z= $4Y_1 - 2Y_2 + Y_3 - 3Y_4$ ii) Find the distribution ii) Find the distribution of $Z_1 = Y_1 + Y_2 + Y_3 + Y_4$ and $Z_2 = -2Y_1 + 3Y_2 + Y_3 - Y_3 + Y_4$ $2Y_4$

- Compare LDA with Other Dimensionality Reduction Techniques in detail? [CO2] 3-c. 6 What are the applications of Linear discriminant analysis? [CO2] 3-d. 6 Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8). 3.e. 6 Compute the principal component using PCA Algorithm? [CO3] Discuss the image factoring and Least square method? [CO4] 3.f. 6 3.g. Compare the Agnes and Diana algorithm in detail with an example each? [CO5] 6 **SECTION-C** 50 4. Answer any one of the following:-A random sample of size n is drawn from a normal population N($\mu\sigma^2$). Estimate μ 4-a. 10 and σ^2 by the method of maximum likelihood? [CO1]
- If X distributed as $N_3(\mu, \Sigma)$ where ? [CO1] 4-b. 10 μ

	and $\Sigma = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{bmatrix}$	
	Find the P(4X ₁ - 3X ₂ + 5X ₃) < 80	
	Find the $P(5X_2 + 4X_3 > 70)$	
5 Anome	Also define Multivariate normal distribution in detail?	
5. Answer any <u>one</u> of the following		
J-a.	linear and logistic regression? [CO2]	10
5-b.	Differentiate between discriminant analysis and cluster analysis in detail with an example? [CO2]	10
6. Answer any <u>one</u> of the following:-		
6-a.	Discuss the difference and similarity between Supervised learning and dimensionality reduction? [CO3]	10
6-b.	What is the importance of using PCA before the clustering? Discuss the difference between PCA and Clustering? [CO3]	10
7. Answer any <u>one</u> of the following:-		
7-a.	Elaborate the factor analysis model with an example also explain the merits and demerits of using Factor analysis? [CO4]	10
7-b.	Explain Factor Analysis with an example and how does it simplify Research Findings? [CO4]	10
8. Answer any <u>one</u> of the following:-		
8-a.	Discuss the following : 1) Correlations and distances 2) Clustering Profiling 3) Interpreting clusters? [CO5]	10
8-b.	Discuss the Email spam filter problem. To solve this problem what will you choose: Classification or Clustering and why? [CO5]	10
	cop.JUL -	

•