

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA

(An Autonomous Institute Affiliated to AKTU, Lucknow)

B.Tech

SEM: III - THEORY EXAMINATION (2024-2025)

Subject: Computational Statistics

Time: 3 Hours

Max. Marks: 100

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

20

1. Attempt all parts:-

- 1-a. The shape of the normal curve depends on its ____ (CO1,K2) 1
- (a) Mean deviation
 - (b) Quartile deviation
 - (c) Standard deviation
 - (d) correlation
- 1-b. All of the following are examples of dependence methods of analysis EXCEPT? (CO1, K1) 1
- (a) multiple regression analysis
 - (b) multiple discriminant analysis
 - (c) multivariate analysis of variance
 - (d) cluster analysis
- 1-c. Analysis of variance is a statistical method of comparing the _____ of several populations. (CO2,K1) 1
- (a) Standard deviations
 - (b) variances
 - (c) Means
 - (d) Proportions
- 1-d. In linear discriminant function analysis, what is the goal of dimensionality reduction. (CO2,K1) 1

- (a) To increase the number of features
 (b) To decrease the number of features
 (c) To transform into a higher-dimensional space
 (d) To eliminate outliers
- 1-e. The correct statement for t-SNE and PCA is____ (CO3,K1) 1
- (a) t-SNE is linear whereas PCA is non-linear
 (b) t-SNE and PCA both are linear
 (c) t-SNE and PCA both are nonlinear
 (d) t-SNE is nonlinear whereas PCA is linear
- 1-f. Potential drawback of PCA is____ (CO3,K1) 1
- (a) It always improves the interpretability of the data.
 (b) It may not be suitable for nonlinear data.
 (c) It doesn't require any assumptions about the data distribution.
 (d) It cannot handle missing values in the dataset.
- 1-g. If a researcher wants to determine the amount of variance in the original variables that is associated with a factor, s/he would use: (CO4, K2) 1
- (a) Factor loadings
 (b) Communalities
 (c) Eigen values
 (d) Beta coefficients
- 1-h. The following is not the part of the exploratory factor analysis process is (CO4,K1) 1
- (a) Extracting factors
 (b) Determining the number of factors before the analysis
 (c) Rotating the factors
 (d) Refining and interpreting the factors
- 1-i. Sentiment Analysis is an example of: ____ (CO5,K1) 1
- 1.Regression
 2.Classification
 3.Clustering
 4.Reinforcement Learning
- (a) 1 and 3
 (b) 1, 2 and 4
 (c) 1, 2, 3 and 4
 (d) 1 and 2
- 1-j. _____ algorithm is most sensitive to outliers.(CO5, K1) 1
- (a) K-means clustering algorithm
 (b) K-medians clustering algorithm

- (c) K-modes clustering algorithm
- (d) K-medoids clustering algorithm

2. Attempt all parts:-

- 2.a. Given a normal distribution with $x=60$, $\mu = 75$ and $\sigma=9$, find the corresponding z-score. (CO1,K3) 2
- 2.b. Define the concept of ANOVA. (CO1,K1) 2
- 2.c. Write down the significance of coefficient of Principal Component. (CO3,K1) 2
- 2.d. List the assumptions of factor analysis.(CO4,K1) 2
- 2.e. Write down any two Distance measuring clustering methods.(CO5, K1) 2

SECTION-B

30

3. Answer any five of the following:-

- 3-a. Give the likelihood function for Multivariate normal distribution and find out estimate for mean vector and covariance matrix. (CO1,K2) 6
- 3-b. If X distributed as $N_3(\mu, \Sigma)$, where (CO1,K3) 6

$$\mu = \begin{bmatrix} 5 \\ 3 \\ 7 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 9 \end{bmatrix}$$

(a) Find the probability $P(X_1 > 6)$

(b) Find the $P(4X_1 - 3X_2 + 5X_3) < 80$

Note : For Normal distribution $P(0 < z < 0.5) = 0.1915$, $P(0 < z < 0.8) = 0.2881$, $P(0 < z < 1) = 0.3413$, $P(0 < z < 1.5) = 0.4332$, $P(0 < z < 2) = 0.4772$

- 3-c. Explain ANOVA with suitable example and also write the model of two-way classification of ANOVA. (CO2, K2) 6
- 3-d. The following figures relate to the production in kg of three varieties of wheat A, B and C used on 15 plots. (CO2,K3) 6

A	14	17	16	18	21	21
B	15	11	13	20	24	23
C	18	16	18	19	20	17

Test whether there is any significant difference in the production of three varieties or not. ($F_{tab.}=19.43$ at 5%)

- 3.e. Discuss the need of dimensionality reduction.Also write down its drawbacks.(CO3, K2) 6
- 3.f. Explain Factor Loading and its importance.(CO4, K2) 6
- 3.g. Differentiate between the different types of machine learning models. (CO5,K2) 6

SECTION-C

50

4. Answer any one of the following:-

4-a. Define likelihood function for Normal distribution.? (CO1,K3) 10

Let X follows If X distributed as $N_2(\mu, \Sigma)$,

where

$$\mu = 0 \text{ and } \Sigma = \begin{bmatrix} 2 & -1 \\ -1 & 4 \end{bmatrix}$$

Find the distribution of $Z = \begin{bmatrix} X - Y \\ Y \end{bmatrix}$

4-b. A random sample of size n is drawn from a normal population $N(\mu, \sigma^2)$, Estimate μ and σ^2 by the method of maximum likelihood. (CO1, K3) 10

5. Answer any one of the following:-

5-a. Discuss the assumptions made in Linear Discriminant Analysis (LDA) and how violations of these assumptions can affect the performance of the algorithm. (CO2,K1) 10

5-b. Differentiate between linear regression and logistic regression. Also give the practical utility of logistic regression. ? (CO2,K2) 10

6. Answer any one of the following:-

6-a. Differentiate between the PCA and Hierarchical clustering in detail with example.(CO3, K2) 10

6-b. Explain, "How can you evaluate the performance of a dimensionality reduction algorithm on your dataset." with an example. (CO3,K2) 10

7. Answer any one of the following:-

7-a. Discuss the useful way to establish how many factors should be kept in your factor analysis. And also write the applications of SPSS tools.(CO4, K2) 10

7-b. Differentiate between Exploratory Factor analysis and Confirmatory Factor analysis.(CO4,K2) 10

8. Answer any one of the following:-

8-a. Discuss the Email spam filter problem. To solve this problem what will you choose: Classification or Clustering and why? (CO5,K2) 10

8-b. Describe the steps how DBSCAN works. Also,write down the applications of DBSCAN.(CO5,K2) 10