

Total No. of Questions : 6]

SEAT No. :

PE-13069

[Total No. of Pages : 3

[6522]-1001
M.Sc. (Data Science)
DS-501-MJ : Statistics for Data Science
(2024 Pattern) (Semester - I)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates :

- 1) *All questions are compulsory Q.1 to Q. 6.*
- 2) *Figures to the right indicate full marks.*
- 3) *Neat diagrams must be drawn wherever necessary.*
- 4) *Use of Statistical tables and scientific calculator is allowed.*

Q1) Attempt any Six of the following :

[6 × 2 = 12]

- a) Define statistics and explain its role in data science.
- b) Calculate the coefficient of variation for the following data :
Mean = 50, S.D = 10.
- c) What is the difference between positive correlation and negative correlation?
- d) State the limits of Karl Pearson's coefficient of correlation and interpret $r = -0.85$.
- e) Define probability using the classical approach.
- f) Define the coefficient of determination (r^2) and explain its significance in regression analysis.
- g) Differentiate between Type I and Type II errors in hypothesis testing.

Q2) Attempt any Four of the following :

[4 × 3 = 12]

- a) Calculate the variance and standard deviation for the following data : 12, 15, 18, 20, 25.
- b) What is a scatter diagram? How is it useful in studying correlation between two variables?
- c) Explain the concept of multiple correlation coefficient. How is it different from simple correlation coefficient?
- d) State the multiplication theorem of probability. If $P(A) = 0.6$, $P(B) = 0.5$, and $P(A \cap B) = 0.3$, find $P(A|B)$ and $P(B|A)$.
- e) What is the difference between linear regression and logistic regression? Give one application of each.

P.T.O.

Q3) Attempt any Three of the following :**[3 × 4 = 12]**

- a) For the following data, calculate the Median, and Mode :

Monthly income (in ₹ 000) of 40 families

Class Interval	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	12	10	5

- b) Explain the measures of skewness and kurtosis. How are they useful in understanding the distribution of data?
- c) Explain the identification and solution methods for multicollinearity in multiple regression analysis. Why is it a problem?
- d) Write a Note on Stratified Random Sampling.

Q4) Attempt any Three of the following :**[3 × 4 = 12]**

- a) Calculate Karl Pearson's coefficient of correlation for the following data :

Temperature in °C (X) and ice cream sales in units (Y)

X	10	12	15	18	20
Y	25	28	30	35	40

Also interpret the result.

- b) Fit a linear regression line $Y = a + bX$ for the following data:

Study hours (X) and exam scores (Y) of 5 students

X	1	2	3	4	5
Y	2	5	7	10	12

- c) A bag contains 5 red and 3 black balls. Two balls are drawn without replacement. Find the probability that i. Both are red ii. Both are black
- d) Two random samples of sizes 100 and 150 are drawn from two populations with means 68 and 65 and standard deviations 3 and 4 respectively. Test at 5% level of significance whether the difference between sample means is significant.

Q5) Attempt the following questions :

a) Attempt any Two of the following : [2 × 5 = 10]

- i) Calculate Bowley's coefficient of skewness for the following data :
Age distribution of 65 students in a college.

Class Interval	10-20	20-30	30-40	40-50	50-60	60-70
Frequency	4	8	15	20	12	6

- ii) Write a note on SRSWR and SRSWOR
- iii) A random sample of 100 students has a mean score of 72 with a standard deviation of 8. Test the hypothesis that the population mean is 70 at 5% level of significance.

b) Attempt any One of the following : [1 × 2 = 2]

- i) What is data sanitization? Why is it important in data analysis?
- ii) Define Correlation.

Q6) Attempt any Two of the following :

[2 × 5 = 10]

- a) Explain the procedure of Chi-square test for independence of attributes for the following data : A survey was conducted to study the relationship between gender and preference for a new product. The following data was obtained :

Gender	Like	Dislike	Total
Male	120	80	200
Female	150	100	250
Total	270	180	450

Test at 5% level of significance whether there is any association between gender and product preference.

- b) Define Simple Linear Regression. Derive the formulas for regression coefficients 'a' and 'b' in the equation $Y = a + bX$.
- c) Explain the procedure of Chi-square test for goodness of fit following data :
A die is thrown 120 times with the following results:

Number on die	1	2	3	4	5	6
Frequency	25	17	15	23	24	16

Test whether the die is unbiased at 5% level of significance using Chi-square test.



Total No. of Questions : 7]

SEAT No. :

PE-13070

[Total No. of Pages : 3

[6522]-1002

F.Y. M.Sc. (Data Science)

DS-502-MJ : COMPUTATIONAL MATHEMATICS

(2024 Pattern) (Semester - I)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

- 1) All questions are compulsory.
- 2) Figures to the right indicate full marks.
- 3) Use of single memory, non-programmable scientific calculator is allowed.

Q1) Attempt any five of the following :

[5 × 2 = 10]

a) Define vector space with one example.

b) Define Linear map with example.

c) Compute inverse of matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$

d) Let $A = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 2 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$, $C = [2, -2]$ Compute ABC.

e) Determine whether the given set of vectors orthogonal or not.

$$u = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, v = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$$

f) Find characteristic polynomial of matrix.

$$A = \begin{bmatrix} 3 & 1 \\ 0 & -2 \end{bmatrix}$$

g) Define Basis and Dimension.

P.T.O.

Q2) Attempt any two of the following :

[2 × 5 = 10]

- a) Prove that the intersection of two subspace is also a subspace.
- b) Find the eigenvalue and eigenvectors of matrix

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

- c) Apply Gram-Schmidt process to vectors (2, 0), (3, 5)

Q3) Attempt any two of the following :

[2 × 5 = 10]

- a) Define Injective, Surjective and Bijective linear maps with examples.
- b) Find whether the vectors (2, -1, 4), (3, 6, 2), (2, 10, -4) are Linearly independent or not.
- c) Define orthogonality of vector. Check whether the vectors (1, 2, 2) and (2, -1, 0) are orthogonal.

Q4) Attempt any two of the following :

[2 × 5 = 10]

- a) Explain QR Decomposition of a matrix with steps.
- b) Represent the linear transformation $T(x, y) = (x + 2y, 3x + y)$ as a matrix and find its rank.
- c) Find the basis and dimension of the subspace spanned by
 $S = \{(1, 0, 1), (0, 1, 1), (1, 1, 2)\}$

Q5) Attempt any two of the following :

[2 × 5 = 10]

- a) Explain Least Squares method with a small example.
- b) What is inner product space? State any two properties. Find the norm of vector $V = (2, -3, 6)$
- c) Verify whether the map $T(x, y, z) = (x + y, y + z)$ is linear or not.

Q6) Attempt any two of the following :

[2 × 5 = 10]

- a) Define Singular Value Decomposition (SVD) and state it's applications.
- b) Solve the system using Row Echelon Form :
 $x + y + z = 6, 2x + 3y + z = 10, x + 2y + 3z = 13$
- c) Show that eigenvalues of an upper triangular matrix are it's diagonal elements.

Q7) Attempt any one of the following.

[1 × 10 = 10]

- a) Find the basis for the row space, column space & Null space of A.
where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 1 & 1 \end{bmatrix}$$

verify the rank - nullity theorem.

- b)
 - i) Explain SVD in Handwritten Digit Recognition and discuss it's role in Dimensionality Reduction.
 - ii) Find a matrix P that diagonalize $A = \begin{bmatrix} 1 & 4 \\ 1 & -2 \end{bmatrix}$ and determine $P^{-1}AP$.



Total No. of Questions : 5]

SEAT No. :

PE-13071

[Total No. of Pages : 2

[6522]-1003

F.Y. M.Sc. (Data Science)

**DS-503-MJ : FUNDAMENTALS OF DATA SCIENCE
(2024 Pattern) (Semester - I)**

Time : 2 Hours]

[Max. Marks : 35

Instructions to the candidates:

- 1) *Question 1 is compulsory.*
- 2) *Answer any 3 questions from Q2 to Q5.*
- 3) *Question No. 2 to 5 carry equal marks.*

Q1) Attempt any five of the following :

[5 × 1 = 5]

- a) Define Open Data Source.
- b) What is tag cloud?
- c) Define statistical data analysis.
- d) Define Data Discretization.
- e) What do you mean by Noisy data?
- f) What is geographical data?

Q2) Attempt the following :

- a) List data visualization libraries in Python. **[2]**
- b) What is 3V's? **[4]**
- c) What do you mean by Standard Datasets? List name of two standard datasets? **[4]**

Q3) Attempt the following :

- a) What do you mean by Standardization. **[2]**
- b) What do you mean by hypothesis testing? Explain parametric and non-parametric hypothesis test. **[4]**
- c) Explain different types of data attributes with example. **[4]**

P.T.O.

Q4) Attempt the following :

- a) Define data reduction. [2]
- b) What is outlier? List types of outlier and explain methods of outlier detection. [4]
- c) Explain data science life cycle. [4]

Q5) Attempt any two of the following : [2 × 5 = 10]

- a) What do you mean by dendrogram? Explain use of dendrogram in brief. [5]
- b) What are different types of data. Explain with examples. [5]
- c) Differentiate between feature selection and feature extraction. [5]



Total No. of Questions : 5]

SEAT No. :

PE13072

[Total No. of Pages : 2

[6522]-1004

First Year M.Sc. (Data Science)

**DS-510-MJ : DATA MINING AND DATA WAREHOUSING
(2024 Pattern) (Semester-I)**

Time : 2 Hours]

[Max. Marks : 35

Instructions to the candidates:

- 1) *Q. 1 is compulsory.*
- 2) *Attempt any 3 questions from 2 to 5.*
- 3) *Figures to the right indicate full marks.*

Q1) Attempt any 5 of the following:

[5×1=5]

- a) What is Data mart?
- b) Define Discretization.
- c) What is OLAP?
- d) What is Metadata?
- e) Define clustering.
- f) Explain Association Rule.
- g) What is KDD?

Q2) Attempt the following:

[2+4+4=10]

- a) Explain Data warehouse architecture with a neat diagram.
- b) Explain data preprocessing techniques in detail.
- c) Explain apriori algorithm for finding frequent itemsets.

P.T.O.

Q3) Attempt the following:

[2+4+4=10]

- a) What is data mining? Explain its applications.
- b) What are the measures for selecting best split to construct a decision tree?
- c) Explain various OLAP operations with example.

Q4) Attempt the following:

[10]

- a) Explain the role of data warehousing in business analytics.
- b) What are data visualization techniques used in data mining?
- c) Write short note on multilevel association rule.

Q5) Attempt any 2 of the following

[5+5=10]

- a) Explain K-Means clustering algorithm with advantages and limitations.
- b) Explain naive -bayes classifier with its advantages & disadvantages.
- c) What are query tools? Explain the applications of query tools in data warehousing.



Total No. of Questions :5]

SEAT No. :

PE13073

[Total No. of Pages : 2

[6522]-1005

F.Y. M.Sc. (Data Science)

DS-512-MJ : ARTIFICIAL INTELLIGENCE

(2024 Pattern) (Semester-I)

Time : 2 Hours]

[Max. Marks : 35

Instructions to the candidates:

- 1) *Q.1 is compulsory.*
- 2) *Solve any three questions from Q.2 to Q.5.*
- 3) *Figures to the right indicate full marks.*

Q1) Solve any five of the following.

[5×1=5]

- a) What is an intelligent agent?
- b) Define state space.
- c) Define Monte-carlo tree search.
- d) List any two Informed search techniques.
- e) What is heuristic search?
- f) Define knowledge representation.

Q2) Attempt the following.

[10]

- a) What are the applications of Artificial Intelligence? **[2]**
- b) Explain water Jug Problem in details. **[4]**
- c) Explain Breadth First Search (BFS). **[4]**

P.T.O.

Q3) Attempt the following. [10]

- a) Define and structure of intelligent agent. [2]
- b) Explain the role of AI in Data science. [4]
- c) Explain A* Algorithm in details. [4]

Q4) Attempt the following. [10]

- a) Explain the techniques of knowledge representation. [2]
- b) Construct the truth table for [4]
 - i) $p \vee q \sim q \rightarrow p$
 - ii) $(\sim (p \wedge q) \wedge r) \rightarrow \sim p$
- c) Role of Artificial Intelligence in Data Science. [4]

Q5) Attempt the following. [10]

- a) Describe AI knowledge cycle. [2]
- b) Explain Tower of Hanoi and Travelling salesman problem. [4]
- c) Explain Propositional Logic and Predicate logic. [4]



Total No. of Questions : 6]

SEAT No. :

PE13074

[Total No. of Pages : 2

[6522]-1006

First Year M.Sc. (Data Science)

DS - 531-RM: RESEARCH METHODOLOGY

(2024 Credit Pattern) (Semester - I)

Time : 3 Hours]

[Max. Marks : 70

Instructions to the candidates:

- 1) *All questions are compulsory.*
- 2) *Figures to the right indicate full marks.*
- 3) *Neat diagrams must be drawn wherever necessary.*

Q1) Solve any Six of the following.

[12]

- a) State two objective of research.
- b) What is exploratory research?
- c) State one method of primary data collection.
- d) Define publication ethics.
- e) Define research problem.
- f) What is chi-square test?
- g) What is predatory journal?
- h) What is measure of central tendency?

Q2) Answer any four of the following.

[12]

- a) Distinguish between research methods and research methodology.
- b) Explain the need and importance of ethical standards in academic publishing.
- c) What are the essential characteristics of good research? Explain any five.
- d) What is research designs and explain different types of research designs.
- e) Explain techniques of interpretation.

P.T.O.

Q3) Answer any three of the following. [12]

- a) Explain meaning, objectives and motivation of research.
- b) What is Sampling? Explain types of Sampling.
- c) Differentiate between survey research and case study method.
- d) Explain journal finder and plagiarism detection tools.

Q4) Answer any three of the following. [12]

- a) Explain sources of literature.
- b) Explain steps in research process.
- c) What is a hypothesis? Explain different types of hypothesis.
- d) Explain the role of researcher and the qualities required to conduct good research.

Q5) Attempt the following:

a) Solve any two of the following. [2×3=6]

- i) Discuss t-test.
- ii) Explain formulation, selection of a research problem.
- iii) Discuss data processing and analysis techniques.

b) Solve any two of the following. [2×3=6]

- i) What is literature review?
- ii) Explain are the qualities of good research?
- iii) State the issues in authorship?

Q6) Answer any two of the following. [10]

- a) What is primary data? Explain the methods of primary data collection.
- b) Explain experimental research design in detail.
- c) What is research report? Mention any two types of research reports.

