

Roll No.

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

Total No. of Pages : 03

Total No. of Questions : 10

MBA (Sem.-3)
DATA SCIENCES USING R

Subject Code : MBA-962-18

M.Code : 77058

Date of Examination : 22-12-2024

Time : 3 Hrs.

Max. Marks : 60

INSTRUCTIONS TO CANDIDATES :

1. SECTION-A contains EIGHT questions carrying TWO marks each and students has to attempt ALL questions.
2. SECTION-B consists of FOUR Subsections : Units-I, II, III & IV. Each Subsection contains TWO questions each carrying EIGHT marks each and student has to attempt any ONE question from each Subsection.
3. SECTION-C is COMPULSORY and consist of ONE question carrying TWELVE marks.

SECTION-A

1. Explain briefly:

- a) Data pre-processing.
- b) Describe any two R commands.
- c) Bayes' Theorem.
- d) Decision Tree.
- e) Agglomerative Clustering.
- f) Bagging and Boosting.
- g) Cross validation.
- h) Data Mining.

SECTION-B

UNIT-I

2. Discuss different components of data science. Also highlight the roles and responsibilities of a data scientist.

3. What is R programming language? Explain how standard deviation, variance, correlation can be found through R? Do you think R-charts can be helpful here?

UNIT-II

4. Decision tree is a simple classification tool, justify your answer with the help of an example.
5. What do you understand by regression? How linear and multiple regressions are different from logistic regression?

UNIT-III

6. What are Ensemble1 methods? Discuss any two such methods.
7. What is clustering? Discuss various methods of clustering.

UNIT-IV

8. How cross-validation and holdout can be used to estimate performance of a classifier.
9. Discuss about big data related techniques.

SECTION-C

10. **Case Study :**

You have the e-commerce sales data against the online advertising expenditure for five online stores in Table-1.

Online Store	e-commerce sale-(Y)	Advertisement Expenditure (X)
S1	36	1.7
S2	34	1.5
S3	65	2.8
S4	95	5
S5	55	2.5

Table-1 (All Figures in Lakhs)

Questions :

- a) Find the dependent and independent variables.
- b) Find the equation of the straight line that fits the data set (linear regression) $Y = A + BX$.
- i) Where A is the expected sale with no Adv-expenditure.

- ii) Where B is average change in total sale with unit change in Adv-expenditure.
 - c) Find the expected sales, if adv-expenditure is 3 lakhs.
 - d) List R command to find mean, variance and covariance.
- (Hint: to find values of A and B you need to find mean and covariance of X and Y)

NOTE : Disclosure of Identity by writing Mobile No. or Making of passing request on any page of Answer Sheet will lead to UMC against the Student.