

THE KAVERY ENGINEERING COLLEGE*(An Autonomous Institution)***B.E/B.TECH. DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL-MAY 2025****Fourth Semester****Artificial Intelligence and Data Science****AD3491 - Fundamentals of Data Science and Analytics****Regulations – 2021****Duration : 3 Hrs****Maximum : 100 Marks****PART – A (ANSWER ALL QUESTIONS) (Marks 10*2=20)**

Q.No	Questions	BLT	CO	Marks
1.	Enlist any four benefits of data science.	L1	1	2
2.	State the need of confusion matrix.	L2	1	2
3.	What are called outliers?	L1	2	2
4.	Where scatter plots are used? Give an example.	L2	2	2
5.	Compare one-tailed and two-tailed tests.	L2	3	2
6.	State central limit theorem.	L2	3	2
7.	Define ANOVA.	L1	4	2
8.	Where will you perform f-test?	L2	4	2
9.	Specify the importance of exponentially weighted moving average.	L2	5	2
10.	What is survival analysis?	L1	5	2

PART – B (ANSWER ALL QUESTIONS) (Marks 5*16 = 80)

Q.No	Questions	BLT	CO	Marks
11.	a i Brief the six steps in the data science process.	L2	1	8
	ii Elaborate the different facets of data.	L2	1	8
	Or			
	b Explain about cleansing, integrating, and transforming data with suitable examples.	L2	1	16
12.	a i Predicted card returns (Y ') for different card investments (X) is given below:	L3	2	8
	X Y '			
	0 6.40			
	4 9.60			
	8 12.80			
	10 14.40			
	12 16.00			
	20 22.40			
	30 30.40			
	Calculate the standard error of estimate for the data available, assuming that the correlation of .30 is based on n = 35 pairs of observations.			

- ii Determine the values of the range and the IQR for the following sets of data. L3 2 8
- (a) Retirement ages: 60, 63, 45, 63, 65, 70, 55, 63, 60, 65, 63
- (b) Residence changes: 1, 3, 4, 1, 0, 2, 5, 8, 0, 2, 3, 4, 7, 11, 0, 2, 3, 4

Or

- b i Express each of the following scores as a z score: L3 2 8
- (a) Margaret's IQ of 135, given a mean of 100 and a standard deviation of 15
- (b) A score of 470 on the SAT math test, given a mean of 500 and a standard deviation of 100
- (c) A daily production of 2100 loaves of bread by a bakery, given a mean of 2180 and a standard deviation of 50

- ii The scores of 30 students in a test are grouped into the following frequency distribution: L3 2 8

Class Interval Frequency (f)

0 - 10	2
10 - 20	4
20 - 30	6
30 - 40	10
40 - 50	5
50 - 60	3

Find the mean of the distribution, the standard deviation and the coefficient of variation.

13. a i A school district is evaluating math proficiency among eighth graders. A score of 8.0 indicates proficiency appropriate for the 8th-grade level. Scores below 8.0 suggest underperformance, and scores above 8.0 suggest above-average proficiency. A past study has shown that the population standard deviation of math scores is 0.5. A random sample of 49 eighth graders is selected, and the mean score is found to be 7.85. Construct a 95% confidence interval for the unknown population mean, and interpret whether the students are underachieving or not. L3 3 8
- ii Explain about estimation method and confidence interval. L2 3 8

Or

- b i A high school teacher is analyzing how many hours students study per week. The weekly study hours for a small group of 5 students are recorded as follows: L3 3 8
- {3, 5, 7, 9, 11}

- List all possible samples of size 2, selected without replacement from this population.
- Construct a relative frequency distribution table showing how often each sample mean occurs.
- Interpret the distribution: Which sample mean(s) appear most frequently? What does this tell you about the distribution of sample means?

- ii Elaborate on null hypothesis test in detail with suitable examples. L2 3 8

14.	a	i	A company currently assumes that the average commute time for its employees is 30 minutes. However, with more employees working remotely or living farther out, management wants to test if this assumed average still holds true. To investigate, they randomly select 8 employees and record their one-way commute times (in minutes): 28, 35, 40, 25, 32, 30, 36, 27 Using a 0.05 level of significance, test the hypothesis that the average commute time is still 30 minutes.	L3	4	8
		ii	Explain about one-factor ANOVA in detail.	L2	4	8
Or						
	b	i	Blood pressure of 8 patients are before and after are recorded: Before: 180, 200, 230, 240, 170, 190, 200 and 165 After: 140, 145, 150, 155, 120, 130, 140 and 130. Find, is there any significant difference between BP reading before and after by applying two-sample t-test.	L3	4	8
		ii	Explain in detail about the chi-square test with an example.	L2	4	8
15.	a		Explain about logistic regression and multiple regression in detail.	L2	5	16
Or						
	b		Explain about time series analysis with suitable examples.	L2	5	16