SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK

IV SEMESTER

AD3464 FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

Regulation – 2023

Academic Year 2025 – 2026 (EVEN SEMESTER)

Prepared by

R. Vaishnavi, Assistant Professor (O.G)
B. Yogesh Kumar, Assistant Professor (O.G)

SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

QUESTION BANK

SUBJECT: AD3464 - FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

SEM / YEAR: IV SEMESTER/ SECOND YEAR

UNIT 1 INTRODUCTION TO DATA SCIENCE			
Need for data science - benefits and uses - facets of data - data science process - setting the research goal - retrieving data - cleansing, integrating, and transforming data - exploratory data analysis - build the models - presenting and building applications.			
PART – A			
Q.No	**Question**	**BT Level**	**Competence**
1	Mention how Data Science helps in making data-driven decisions.	BTL 2	Understand
2	What is Bigdata?	BTL 2	Understand
3	What is machine Learning?	BTL 1	Remember
4	Define Data Mining?	BTL 1	Remember
5	List the major sources from which big data is generated.	BTL 1	Remember
6	Mention the categories of data.	BTL 2	Understand
7	List some of the application domains of data science.	BTL 1	Remember
8	What is structured data? Give some examples.	BTL 1	Remember
9	structured data with suitable examples.	BTL 1	Remember
10	why machine-generated data is crucial in modern analytics.	BTL 2	Understand
11	Why the data is to be cleaned.	BTL 2	Understand
12	List the phases involve in the data science process.	BTL 1	Remember
13	What is meant by data cleaning?	BTL 1	Understand
14	Discuss the concept and purpose of a project charter.	BTL 2	Remember
15	Identify the important contents of a project charter.	BTL 1	Remember

16	List some of the visualization techniques	BTL 2	Understand
17	What are all the challenges faced while collecting real-time data.	BTL 2	Understand
18	Define data warehouse, DataMart and data lake.	BTL 2	Understand
19	List some of the factors involved in selecting the modeling technique.	BTL 2	Understand
20	What is a dummy variable?	BTL 1	Remember
21	What do you meant by exploratory data analysis?	BTL 1	Remember
22	List out the methods for combining data from different table.	BTL 1	Remember
23	why predictive modeling is essential in business applications.	BTL 2	Understand
24	On what factors the modeling technique is being selected.	BTL 2	Understand

PART – B

Q.No	Question	BT Level	Competence
1	Discuss the applications of data science and bigdata with suitable examples.	BTL 6	Create
2	Illustrate the overview of the data science process.	BTL 4	Analyze
3	Evaluate the role of data science in five current technological fields and justify why it is essential in each field.	BTL 5	Evaluate
4	Describe the categories of data for data mining.	BTL 3	Apply
5	Analyze how setting a clear research aim influences the success and direction of a data science project.	BTL 4	Analyze
6	Discuss the categories involved in retrieving relevant data from different sources of data.	BTL 5	Evaluate
7	Develop a detailed workflow for preparing raw datasets for analysis, including all essential processing steps.	BTL 6	Create
8	Elucidate the techniques involved in data cleansing.	BTL 4	Analyze
9	Illustrate the steps involved in combining data from different data sources.	BTL 6	Create
10	Propose a framework showing how automated data generation can be integrated into a large-scale analytics system to enhance insights.	BTL 6	Create
11	Elaborate on the steps involve in model building with suitable diagrams.	BTL 3	Apply
12	Discuss briefly about facets of data.		
13	Justify Exploratory Data Analysis.	BTL 4	Analyze
14	Explain briefly about Data science and its life cycle.		
15	Analyze how Data Science differs from Big Data in terms of purpose, processing techniques, and outcomes.	BTL 4	Analyze
16	Compare and contrast Cloud Computing and Big Data.	BTL 4	Analyze
17	Explain the impact of Big Data technologies on the field of Data Science. How do these technologies enhance the capacity of Data Scientists to solve complex problems?	BTL 3	Apply

UNIT 2 DESCRIPTIVE ANALYTICS

Frequency distributions - Outliers - interpreting distributions – graphs - averages – describing variability - interquartile range - variability for qualitative and ranked data - Normal distributions - z scores – correlation - scatter plots – regression - regression line - least squares regression line - standard error of estimate - interpretation of r2 - multiple regression equations - regression toward the mean.

PART – A

Q.No	Question	BT Level	Competence
1	What is meant by frequency distribution?	BTL 1	Remember
2	What is non-numerical (category-based) data with suitable examples.	BTL 1	Remember
3	What is meant by quantitative data? Give examples.	BTL 1	Remember
4	Differentiate qualitative and quantitative data	BTL 1	Remember
5	Compare discrete and continuous variables.	BTL 1	Remember
6	State the difference between nominal and ordinal data	BTL 1	Remember
7	Mention the types of frequency distribution?	BTL 1	Remember
8	Define an outlier?	BTL 2	Understand
9	What is percentile rank?	BTL 1	Remember
10	Provide the equation for percentile rank.	BTL 2	Understand
11	Differentiate between histograms and bar charts in representing information.	BTL 2	Understand
12	Give the measures of central tendency	BTL 2	Understand
13	What do you call the value that appears most frequently in a dataset?	BTL 1	Remember
14	Define median.	BTL 1	Remember
15	What is the interpretation of r^2 ?	BTL 1	Remember
16	What is the standard error of estimate?	BTL 2	Understand
17	What term is used to describe the average amount by which data values differ from their mean?	BTL 1	Remember
18	What is normal curve?	BTL 2	Understand
19	Define z-score.	BTL 2	Understand
20	Give the equation for z-score.	BTL 1	Remember
21	How can a standardized z-score be transformed back into its corresponding raw score?	BTL 1	Remember
22	Define Correlation.	BTL 1	Remember
23	Mention the types of correlation.	BTL 2	Understand
24	Define Scatterplot	BTL 2	Understand

PART – B

<table>
<tr><th>Q.No</th><th>Question</th><th>BT Level</th><th>Competence</th></tr>
<tr><td>1</td><td>Explain the different types of frequency distribution with suitable examples and diagrams.</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>2</td><td>Elaborate the different ways to describe or represent data using tables with suitable examples.</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>3</td><td>Analyze the role of graphical methods in presenting statistical data, and explain how different types of graphs can highlight various patterns within a dataset.</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>4</td><td>Compute the mean, median and mode for the following datasets
.I)9,10,12,13,13,13,15,15,16,16,18,22,23,24,24,25</td><td>BTL 3</td><td>Apply</td></tr>
<tr><td>5</td><td>The following data are the shoe sizes of 50 male students. The sizes are discrete data since shoe size is measured in whole and half units only. Construct a histogram and calculate the width of each bar or class interval. Suppose you choose six bars.
9;9;9.5;9.5;10;10;10;10;10;10;10.5;10.5;10.5;10.5;10.5;10.5;10.5;10.5,
11;11;11;11;11;11;11;11;11;11;11;11;11.5;11.5;11.5;11.5;11.5;11.5
;11.5,12;12;12;12;12;12;12;12.5;12.5;12.5;12.5;14</td><td>BTL 3</td><td>Apply</td></tr>
<tr><td>6</td><td>Evaluate the usefulness of scatterplots in identifying relationships between variables, and compare the different forms such visualizations can take.</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>7</td><td>Elaborate on the correlation coefficient. Compare the various correlation coefficients</td><td></td><td></td></tr>
<tr><td>8</td><td>Explain the characteristics of a normal distribution. Discuss why the normal distribution is widely used in statistics and how it relates to other probability distributions. How can you check if a dataset approximates a normal distribution?</td><td>BTL 6</td><td>Create</td></tr>
<tr><td>9</td><td>What is a z-score, and how is it used to standardize data in statistical analysis? Discuss its role in comparing data points from different distributions and how it helps in identifying outliers.</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>10</td><td>Evaluate how the coefficient of determination (r^2) reflects the performance of a regression model and describe what it indicates about the strength of association between predictors and outcomes.</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>11</td><td>Find Karl Pearson’s Correlation Coefficient for the following paired data.
<table>
<tr><td>X</td><td>38</td><td>45</td><td>46</td><td>38</td><td>35</td><td>38</td><td>46</td><td>32</td><td>36</td><td>38</td></tr>
<tr><td>Y</td><td>28</td><td>34</td><td>38</td><td>34</td><td>36</td><td>36</td><td>28</td><td>29</td><td>25</td><td>26</td></tr>
</table></td><td>BTL 6</td><td>Create</td></tr>
<tr><td>12</td><td>Discuss Multiple Regression Equations.</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>13</td><td>A random sample of 5 college students is selected and their grades in operating system and software engineering are found to be?
<table>
<tr><td>Subject</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Operating System</td><td>85</td><td>60</td><td>73</td><td>40</td><td>90</td></tr>
<tr><td>Software Engineering</td><td>93</td><td>75</td><td>65</td><td>50</td><td>80</td></tr>
</table>
Calculate Pearson’s rank correlation coefficient?</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>14</td><td>Analyze how quartiles help summarize the spread of a dataset, and explain how the interquartile range (IQR) can be applied to detect unusually high or low observations.</td><td>BTL 4</td><td>Analyze</td></tr>
</table>

15	Explain how measures of variability (such as range, variance, and standard deviation) describe the spread of data. Provide examples of when each measure is most appropriate to use.	BTL 5	Evaluate
16	Discuss the phenomenon of regression toward the mean and how it can influence statistical interpretation. Discuss how this concept is related to the correlation between variables and its implications in predictive modeling.	BTL 6	Create
17	Analyze the process involved in conducting simple linear regression and discuss why the regression line is essential for interpreting variable relationships.	BTL 4	Analyze

UNIT 3 - INFERENTIAL STATISTICS

Populations – samples - random sampling - Sampling distribution - standard error of the mean - Hypothesis testing - z-test - z-test procedure - decision rule – calculations - decisions - interpretations - one-tailed and two-tailed tests – Estimation - point estimate - confidence interval - level of confidence - effect of sample size.

PART – A

Q.No	Question	BT Level	Competence
1.	Define population? Give an example.	BTL 1	Remember
2	why real population is important for accurate statistical conclusions?	BTL 2	Understand
3	List the different types of population.	BTL 1	Remember
4	What is hypothetical population?	BTL 1	Remember
5	Define Samples.	BTL 1	Remember
6	List the categories of sample.	BTL 2	Understand
7	What does it mean when a sample is selected such that every member has an equal chance of being chosen?	BTL 1	Remember
8	Mention the types of random sampling.	BTL 1	Remember
9	Differentiate population and sample.	BTL 1	Remember
10	List the types of non-probability sampling.	BTL 2	Understand
11	What is the idea behind snowball sampling and explain how participants help in expanding the sample?	BTL 2	Understand
12	Differentiate non-probability and probability sampling.	BTL 1	Remember
13	Give the optimal sample size.	BTL 2	Understand
14	What is meant by selecting every k-th element from a list during sampling?	BTL 1	Remember
15	Define cluster sampling.	BTL 1	Remember
16	Mention the advantages of random sampling.	BTL 2	Understand
17	Define consecutive sampling.	BTL 1	Remember
18	Provide the standard error of the mean	BTL 1	Remember
19	Give the level of confidence.	BTL 2	Understand
20	State the difference between tests that check for directional effects and	BTL 1	Remember

	those that check for effects in both directions.		
21	Define estimation.	BTL 1	Remember
22	What are the possible decisions you can make after performing a hypothesis test?	BTL 1	Remember
23	Describe the basic steps involved in conducting a Z-test.	BTL 2	Understand
24	State the importance of random sampling in statistical analysis.	BTL 2	Understand

PART – B

Q.No	Question	BT Level	Competence
1	Discuss on population and samples with suitable examples.	BTL 4	Analyze
2	Design and propose a real-world scenario where you would use different random sampling methods to select participants for a study.	BTL 6	Create
3	Elaborate on the different types of non-probability-based sampling techniques.	BTL 5	Evaluate
4	Illustrate the hypothesis testing with an example.	BTL 6	Create
5	Critically evaluate the steps involved in performing a z-test by applying them to a real-life numerical example.	BTL 5	Evaluate
6	Explain in detail about Estimation and the significance of point estimates.	BTL 5	Evaluate
7	Elaborate on Confidence interval and level of confidence.	BTL 6	Create
8	Discuss z-Test Problem.	BTL 4	Analyze
9	Evaluate how decision rules influence outcomes in hypothesis testing, and discuss their role in avoiding incorrect statistical conclusions.	BTL 5	Evaluate
10	What is data interpretation? Discuss Qualitative and Quantitative Data Interpretation.	BTL 3	Apply
11	Analyze how increasing or decreasing the sample size affects the accuracy, variability, and reliability of statistical estimates.	BTL 4	Analyze
12	Find the standard error of mean of given observations, x=10,20,30,40,50	BTL 3	Apply
13	Compare and Contrast one-tailed test and a two-tailed test in hypothesis testing.	BTL 4	Analyze
14	Analyze how the sampling distribution behaves when sample size increases, and explain why larger samples tend to produce more stable mean values.	BTL 4	Analyze
15	Explain how sample size influences the width of a confidence interval. Why does increasing the sample size lead to a more precise estimate?	BTL 5	Evaluate
16	Describe the procedure for conducting a z-test. Outline the steps involved in performing a z-test, from formulating hypotheses to making a decision.	BTL 6	Create
17	Elaborate the steps to test a hypothesis.	BTL 4	Analyze

UNIT 4 - ANALYSIS OF VARIANCE

t-test for one sample - sampling distribution of t - t-test procedure - t-test for two independent samples - p-value - statistical significance - t-test for two related samples. F - test – ANOVA – Two-factor experiments - three f-tests - two-factor ANOVA - Introduction to chi-square tests.

PART – A

Q.No	Question	BT Level	Competence
1	Define categorical variable. Give example.	BTL 1	Remember
2	Mention the types of categorical variable	BTL 2	Understand
3	Mention the major categories into which populations are classified.	BTL 1	Remember
4	What is t -test?	BTL 1	Remember
5	Give the measures of the t-test	BTL 2	Understand
6	When to use the t-test?	BTL 2	Understand
7	Define random sampling and state when it is used in selecting participants.	BTL 1	Remember
8	Can the t-test is used to measure the difference among several groups.	BTL 2	Understand
9	Define chi-square test and write its formulae.	BTL 1	Remember
10	Specify the purpose of chi-square test.	BTL 2	Understand
11	How the chi-square test is interpreted.	BTL 2	Understand
12	What is an acceptable value in chi-square method	BTL 2	Understand
13	Define f-test.	BTL 1	Remember
14	Define systematic sampling and give one simple example of its procedure.	BTL 1	Remember
15	Give the critical value for the F-test.	BTL 1	Remember
16	Why does Anova uses F-test?	BTL 2	Understand
17	Is it possible for a negative F-statistic in a F-test.	BTL 1	Remember
18	How F-test is differentiated from T.	BTL 2	Understand
19	what is meant by the level of confidence in statistical estimation.	BTL 2	Understand
20	How Anova's statistical significance is determined.	BTL 1	Remember
21	What is factorial anova?	BTL 1	Remember
22	State the decisions that can be taken once a hypothesis test is completed.	BTL 1	Remember
23	What is meant by P-Value?	BTL 2	Understand
24	How is P-Value Calculated?	BTL 2	Understand

PART – B

Q.No	Question	BT Level	Competence
1	Elaborate T-test Problem and theory.	BTL 5	Evaluate
2	Analyze how an F-test can be used to compare the variability between two datasets and discuss the reasoning behind the test's decision criteria.	BTL 4	Analyze
3	Analyze how the chi-square test is applied in real-world scenarios and discuss its usefulness in examining relationships between categorical variables.	BTL 4	Analyze

<table>
<tr><td>4</td><td>Discuss ANOVA Problem and theory.</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>5</td><td>Explain briefly about Sampling Distribution of T.</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>6</td><td>a) Illustrate in detail about one factor ANOVA with example. (8)

b) A random sample of 90 college students indicates whether they most desire love, wealth power, health, fame or happiness . Using the .05 level of significance and the following results, test the null hypothesis that in the null
underlying Population, the various desires are equally popular using chi-square test. (8)

<table><tr><th>Freq</th><th>Love</th><th>Wealth</th><th>Power</th><th>Health</th><th>Fame</th><th>Happiness</th><th>Total</th></tr><tr><td>Observed (f0)</td><td>25</td><td>10</td><td>5</td><td>25</td><td>10</td><td>15</td><td>90</td></tr></table></td><td>BTL 6</td><td>Create</td></tr>
<tr><td>8</td><td>A manufacturer of a gas additive claims that it improves gas mileage. A random sample of 30 drivers tests this claim by determining their gas mileage for a full tank of gas that contains the additive (X_1) and for a full tank of gas that does not contain the additive (X_2). The sample mean difference, D, equals 2.12 miles (in favor of the additive), and the estimated standard error equals 1.50 miles.
(i) Using t, test the null hypothesis at the .05 level of significance, (6)
(ii) Specify the p-value for this result. (5)
(iii) Are there any special precautions that should be taken with the present experimental design? (5)</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>9</td><td>Analyze how One-Way ANOVA differs from Two-Way ANOVA in terms of purpose, assumptions, and interpretation, and discuss situations where each method is most appropriately applied.</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>10</td><td>a) A research team wants to study the effects of a new drug on insomnia. 8 tests were conducted with a variance of 600 initially. After 7 months 6 tests were conducted with a variance of 400. At a significance level of 0.05 was there any improvement in the results after 7 months? Evaluate by using f-test. (8)

b) b) Elaborate the difference between F - Test and T – Test. (8)</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>11</td><td>a) A library system lends book for period of 21 days. This policy is being revaluated in view of a possible new loan period that could be either longer or shorter than 21 days. To aid in making this decision, book-lending records were consulted to determine the loan periods actually used by the patrons. A random sample of eight records revealed the following loan periods in days: 21, 15, 12, 24, 20, 21, 13 and 16. Test the null hypothesis with t-test, using the .05 level of significance. (8)

b) Discuss effect size estimation. (8)</td><td>BTL 4</td><td>Analyze</td></tr>
<tr><td>12</td><td>Elaborate F Test in Statistics. Importance of F-Test.</td><td>BTL 5</td><td>Evaluate</td></tr>
<tr><td>13</td><td>Explain the assumptions of the chi-square test and how violations can affect the results.</td><td>BTL 4</td><td>Analyze</td></tr>
</table>

14	Evaluate how the F-distribution helps determine statistical significance in ANOVA and explain why its shape changes with varying degrees of freedom.	BTL 5	Evaluate
15	Derive the formula for the paired t-test statistic.	BTL 4	Analyze
16	Discuss the relationship between the p-value and Type I and Type II errors.	BTL 5	Evaluate
17	Analyze the characteristics of the t-distribution and explain how it differs from the normal distribution in terms of shape and usage.	BTL 4	Analyze

UNIT 5 PREDICTIVE ANALYTICS

Linear least squares - implementation - goodness of fit - testing a linear model - weighted resampling. Regression using Stats Models - multiple regression - nonlinear relationships - logistic regression - estimating parameters - Time series analysis - moving averages - missing values - serial correlation - autocorrelation. Introduction to survival analysis.

PART – A

Q.No	Question	BT Level	Competence
1	State the basic steps involved in computing the least-squares values.	BTL 1	Remember
2	What is meant by “goodness of fit” in statistical analysis?	BTL 2	Understand
3	Define the principle of least square.	BTL 1	Remember
4	Dafne least square.	BTL 1	Remember
5	What is least square curve fitting?	BTL 1	Remember
6	Why is time series analysis essential for identifying long-term patterns and trends in data?	BTL 2	Understand
7	Give some examples for time series analysis.	BTL 1	Remember
8	Mention the types of Time series Analysis	BTL 1	Remember
9	List a few real-world applications where time series analysis is commonly used.	BTL 1	Remember
10	Give the limitations of Timeseries Analysis.	BTL 2	Understand
11	List the Datatypes of Time series.	BTL 2	Understand
12	What does Goodness of fit mean?	BTL 1	Remember
13	Why is Goodness of fit is important?	BTL 2	Understand
14	Provide the most common goodness of fit tests.	BTL 1	Remember
15	Why do we test goodness of fit.	BTL 2	Understand
16	Define multiple linear regression.	BTL 2	Understand
17	How the error is calculated in linear regression model.	BTL 1	Remember
18	What Is Predictive Analytics?	BTL 1	Remember
19	What are the applications of predictive models?	BTL 2	Understand
20	Define Credit.	BTL 1	Remember
21	What is meant by Forecasting?	BTL 1	Remember
22	how predictive analytics differs from machine learning in terms of purpose and application.	BTL 1	Remember

23	Compare Predictive Analytics vs. Machine Learning	BTL 2	Understand
24	Define Regression.	BTL 2	Understand

PART – B

Q.No	Question	BT Level	Competence
1	Explain Multiple regression.	BTL 5	Evaluate
2	Analyze what a survival curve represents and how censoring affects its interpretation.	BTL 4	Analyze
3	Linear least square problem and theory.	BTL 4	Analyze
4	Explain in detail about logistic regression.	BTL 5	Evaluate
5	Evaluate the importance of time series analysis in business forecasting and justify suitable techniques for two selected industries.	BTL 5	Evaluate
6	Write in detail about goodness of fit.	BTL 3	Apply
7	Create a study plan to compare multiple regression and logistic regression using the same dataset.	BTL 6	Create
8	Illustrate in depth about time series forecasting, its components, moving averages and its various methods with examples.	BTL 5	Evaluate
9	Explain in brief about the various steps of Data Analysis.	BTL 3	Apply
10	Describe in detail Introduction to survival analysis.	BTL 4	Analyze
11	Explain in detail serial correlation and autocorrelation.	BTL 4	Analyze
12	Analyze how to build and validate a regression model using stats models, focusing on key diagnostics used for model refinement.	BTL 4	Analyze
13	a) Illustrate nonlinear relationships. (8) b) How to estimate the coefficients using maximum likelihood estimation (MLE) and interpret the estimated parameters. (8)	BTL 5	Evaluate
14	How would you handle missing values and account for them in time series analysis?	BTL 3	Apply
15	Define heteroscedasticity and explain why it is a problem in linear regression.	BTL 3	Apply
16	Apply the concepts of R-squared and adjusted R-squared to explain why adding predictors may lower adjusted R-squared and what decision should follow.	BTL 3	Apply
17	Justify the results of the fitted model, including the significance of the coefficients.	BTL 5	Evaluate