ADITYA COLLEGE OF ENGINEERING & TECHNOLOGY
Survey No. 2/8, 2/9 and 2/2, Kamakshipura, Hesaraghatta Hobli, Yelahanka, Bengaluru, Karnataka, 560089
Approved by AICTE, New Delhi & Affiliated to VTU, Belagavi, Govt. of Karnataka

Semester:7	Department of Data Science	AY: 2026-27
Course Code: BAD702	Course Name: Statistical Machine Learning for Data Science	

QUESTION BANK

Q. No.	QUESTION	M	L	C
	MODULE 1			
1	Differentiate mean, trimmed mean, weighted mean and median with suitable examples.	5	L2	CO1
2	Explain robust estimates of location. Why are they preferred when the data contains outliers?	5	L2	CO1
3	Explain the estimates of variability: variance, standard deviation, mean absolute deviation, median absolute deviation, percentiles and IQR.	5	L2	CO1
4	For the data 12, 15, 11, 18, 20, 14, 95, calculate the mean, median, variance, standard deviation and IQR. Comment on the effect of the outlier.	5	L3	CO1
5	Explain boxplots, frequency tables, histograms and density plots for exploring data distributions.	5	L2	CO1
6	Explain how binary and categorical data are explored using mode, expected value and bar charts.	5	L2	CO1
7	Explain correlation, the correlation matrix and scatterplots. State the limitations of the correlation coefficient.	5	L2	CO1
8	Explain hexagonal binning and contour plots for exploring two numeric variables.	5	L2	CO1
9	Explain how two categorical variables are explored using contingency tables.	5	L2	CO1
10	Explain how violin plots and boxplots are used to compare a numeric variable across categories and to condition on multiple variables.	5	L3	CO1
	MODULE 2			
11	Explain random sampling, sampling bias and selection bias with examples.	5	L2	CO2
12	Explain the sampling distribution of a statistic and the Central Limit Theorem.	5	L2	CO2
13	Define standard error. How does it vary with sample size?	5	L2	CO2
14	Explain the bootstrap procedure with its algorithm and uses.	5	L3	CO2
15	Explain confidence intervals. How is a 90% confidence interval obtained using the bootstrap?	5	L3	CO2
16	Explain the normal distribution, the standard normal distribution, z-scores and the QQ-plot.	5	L2	CO2
17	Explain long-tailed distributions and their importance in data analysis.	5	L2	CO2
18	Explain Student's t-distribution and state when it is used.	5	L2	CO2
19	A call centre receives 4 calls per hour on average. Using the Poisson distribution, find the probability of exactly 6 calls in an hour. Also explain the binomial distribution.	5	L3	CO2
20	Explain the chi-square and F distributions and their applications.	5	L2	CO2
	MODULE 3			
21	Explain A/B testing with an example. Why are a control group and randomization needed?	5	L2	CO3
22	Explain hypothesis testing: null and alternative hypotheses, one-way and two-way tests.	5	L2	CO3
23	Explain the permutation test (resampling) with its steps.	5	L3	CO3
24	Explain statistical significance, p-value, alpha, and Type I and Type II errors.	5	L2	CO3
25	Version A of a webpage got 120 sales from 1,000 users and Version B got 150 sales from 1,200 users. Test at the 5% level whether the conversion rates differ.	5	L3	CO3
26	Explain one-sample, two-sample and paired t-tests with their assumptions.	5	L2	CO3
27	Explain the multiple testing problem. Describe the Bonferroni correction.	5	L2	CO3
28	Explain degrees of freedom and why the divisor n-1 is used for the sample variance.	5	L2	CO3
29	Explain ANOVA and the F-statistic with an example.	5	L2	CO3
30	Explain the chi-square test with an example and discuss Fisher's exact test.	5	L3	CO3

Q. No.	QUESTION	M	L	C
MODULE 4				
31	Explain the multi-arm bandit algorithm. Describe the epsilon-greedy strategy.	5	L2	CO4
32	Explain power, effect size and sample size in statistical experiments.	5	L2	CO4
33	Explain simple linear regression: fitted values, residuals and the least squares method.	5	L2	CO4
34	Explain multiple linear regression and how to interpret the regression equation.	5	L2	CO4
35	Explain factor variables in regression: dummy variables, reference coding and multi-level factors, with an example.	5	L3	CO4
36	Explain regression diagnostics: outliers, influential values and Cook's distance.	5	L2	CO4
37	Explain heteroskedasticity, non-normal residuals and correlated errors, and how residual plots help detect them.	5	L2	CO4
38	Explain partial residual plots and how they detect nonlinearity.	5	L2	CO4
39	Explain polynomial regression with an example.	5	L2	CO4
40	Explain spline regression and generalized additive models.	5	L3	CO4
MODULE 5				
41	Explain the covariance matrix and its role in discriminant analysis.	5	L2	CO5
42	Explain Fisher's linear discriminant with its steps.	5	L3	CO5
43	Explain logistic regression, the logistic response function and the logit.	5	L2	CO5
44	Explain generalized linear models, including the family and link function.	5	L2	CO5
45	Explain how coefficients and odds ratios are interpreted in logistic regression with an example.	5	L3	CO5
46	A Poisson regression model is log(lambda) = 1.8 - 0.02*Age + 0.6*condition. Interpret the coefficients and find the expected visits for a 50-year-old patient with the condition.	5	L3	CO5
47	Explain strategies for imbalanced data: undersampling, oversampling and SMOTE.	5	L2	CO5
48	Explain how a classification model is assessed: confusion matrix, precision, recall, ROC curve and AUC.	5	L2	CO5
49	Differentiate linear discriminant analysis and logistic regression.	5	L2	CO5
50	Explain how class weights and threshold adjustment help with imbalanced data.	5	L3	CO5