

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

April – May 2023	
(B.Tech.) Program: <u>Computer Engineering</u> Scheme : I	
Examination: TY Semester: VI	
Course Code: HDSC601	Course Name: Statistical Learning for Data Science
Date of Exam: 24/05/2023	Duration: 2.5 Hours
Max. Marks: 60	

Instructions:
 (1) All questions are compulsory.
 (2) Draw neat diagrams wherever applicable.
 (3) Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight.	12		
i)	Define Categorical and Quantitative Data.	2	CO1	U
ii)	State the importance of Probability in Data Science.	2	CO2	An
iii)	Explain in brief Population Mean.	2	CO4	U
iv)	State the Null Hypothesis.	2	CO4	U
v)	State the importance of Regression Model.	2	CO5	An
vi)	Describe Sign Test.	2	CO3	U
vii)	What do you mean by Means of Variability.	2	CO2	An
viii)	State the significance of outliers in regard to Data Science.	2	CO3	An
Q.2	Solve any four questions out of six.	16		
i)	Elaborate on Summarizing Quantitative Data and its importance.	4	CO1	U
ii)	According to the National Oceanic and Atmospheric Administration (NOAA), the state of Colorado averages 18 tornadoes every June (NOAA website, November 8, 2012). {note: There are 30 days in June} a. Compute the mean number of tornadoes per day. b. Compute the probability of no tornadoes during a day. c. Compute the probability of exactly one tornado during a day. d. Compute the probability of more than one tornado during a day.	4	CO2	Ap
iii)	A Bag I contains 4 white and 6 black balls while another Bag II contains 4 white and 3 black balls. One ball is drawn at random from one of the bags, and it is found to be black. Find the probability that it was drawn from Bag I.	4	CO2	Ap
iv)	Debrief Systematic Sampling and its purposes with suitable examples.	4	CO3	An

April – May 2023

(B.Tech.) Program: Computer Engineering Scheme: **I**

Examination: TY Semester: VI

Course Name: Statistical Learning for Data Science

Duration: 2.5 Hours

Course Code: HDSC601

Date of Exam: 24/05/2023

Max. Marks: 60

v)	Describe the test procedure for Hypothesis testing.	4	CO4	U
vi)	Explain in brief Outliers and Influential Observations in regards to residual analysis.	4	CO5	U
Q.3	Solve any two questions out of three.	16		
i)	Explain types of Time series patterns with suitable examples and diagrams.	8	CO3	An
ii)	The random variable x is known to be uniformly distributed between 1.0 and 1.5 a. Show the graph of the probability density function. b. Compute $P(x = 1.25)$ c. Compute $P(1.0 \leq x \leq 1.25)$ d. Compute $P(1.20 < x < 1.5)$	8	CO2	Ap
iii)	What are Parametric Tests? Enlist types of Parametric Tests with suitable examples. State advantages and disadvantages of it.	8	CO4	An
Q.4	Solve any two questions out of three.	16		
i)	Elaborate on Cross tabulation and scatter diagram in Statistical Science with suitable examples.	8	CO1	An
ii)	Show the five-number summary and the box plot for the following data: 5, 15, 18, 10, 8, 12, 16, 10, 6.	8	CO3	Ap
iii)	Discuss Simple Linear Regression Model and its utility in data science.	8	CO5	An
