

**Sri Sathya Sai Institute of Higher Learning**

(Deemed to be University)

Vidyagiri, Prasanthi Nilayam

*Master of Science in Data Science and Computing**Final Lab Examination, April 2024***Regression Methods****Duration : 90 mins****02/04/2024****Max. Marks : 25****Answer the following:**

1. For the dataset given below, do the following: (1 + 3 + 2 = 6 Marks)

a. Using R Studio, find the Correlation Coefficient between the dependent and  
the independent variable [1M]

b. Using Excel fit the Simple Linear Regression model on the dataset [3M]

*(Target = #Songs)*

c. Using R Studio, fit the obtained line equation on the Scatter plot of the data [2M]

Months	#Songs
23	486
35	747
2	81
28	581
5	117
32	728
23	445
10	128
4	61
26	476
1	35
8	121
13	266
9	126
5	141

2. For the below given dataset, do the following: (3 + 1 = 4 Marks)

a. Using R Studio, fit the Multiple Linear Regression model on the dataset [3M]

*(Target = GGPA)*

b. Explicitly print the P Values of the two independent columns in R Studio [1M]

GRE Total	UGPA	GGPA
140	3.2	4.0
120	3.7	3.9
125	3.6	3.8
130	2.9	3.7
110	3.5	3.6
100	3.3	3.5
95	3.0	3.4
115	2.7	3.3
105	3.1	3.2
90	2.8	3.1
105	2.4	3.0

3. For the given dataset, do as instructed:

(5 + 2 + 5 + 3 = 15 Marks)

min	Delivery Time, y (min)	Number of Cases, x1	Distance, x2 (ft)
1	16.68	7	560
2	11.5	3	220
3	12.03	3	340
4	14.88	4	80
5	13.75	6	150
6	18.11	7	330
7	8	2	110
8	17.83	7	210
9	79.24	30	1460
10	21.5	5	605
11	40.33	16	688
12	21	10	215
13	13.5	4	255
14	19.75	6	462
15	24	9	448
16	29	10	776
17	15.35	6	200
18	19	7	132
19	9.5	3	36
20	35.1	17	770
21	17.9	10	140
22	52.32	26	810
23	18.75	9	450
24	19.83	8	635
25	10.75	4	150

## MDSC-201

- a. Without using the in-built Data Analysis Toolkit present in Ms-Excel, obtain the MLR equation for the dataset by determining the parameters  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  and  $\sigma^2$  [5M]
- b. Obtain the Hat Matrix H for the given data [2M]
- c. From the Regular Residuals (after fitting the MLR model), derive the following:
  - i. Standardized Residuals [1M]
  - ii. Studentized Residuals [2M]
  - iii. PRESS Residuals [2M]
- d. Using R-Student, create the Normal Probability Plot for the same [3M]

\* \* \* \* \*