4. (a) For the following data, find the regression line of x on y. Also show the regression line in a graph : **5**

x	y
1	9
2	8
3	10
4	12
5	11
6	13
7	14

(b) An industrial designer wants to determine the average amount of time it takes an employee to assemble an "easy-to-assemble' toy. Use the following data (in minutes), a random sample, to construct a 95% confidence interval for the mean of the population sampled : **10**

17	13	18	19	17	21	29	22	16	28	21	15
26	23	24	20	8	17	17	21	32	18	25	22
16	10	20	22	19	14	30	22	12	24	28	11

No. of Printed Pages : 05 **Roll No. .....**

18F115

B.Tech. EXAMINATION, 2022

(Sixth Semester)

(C Scheme) (Main & Re-appear)

(CSE)

CSEH310C

DATA ANALYTICS WITH PYTHON

Time : 3 Hours] [*Maximum Marks : 75*

Before answering the question-paper candidates should ensure that they have been supplied to correct and complete question-paper. No complaint, in this regard, will be entertained after the examination.

Note : Attempt *Five* questions in all. All questions carry equal marks. Students can use distribution tables.

Unit I

1. (a) List and explain the different types of sampling techniques. **5**

(b) Illustrate any *three* measures of central tendency. Also write the pros and cons of each. **10**

2. (a) Find the co-efficient of correlation for the following data : **5**

x	y
5	2
7	2
1	4
3	5
4	6

(b) List and explain the different measures of variability. Also write the pros and cons of each. Use suitable example in support of your answer. **10**

Unit II

3. (a) State and explain the concept of hypothesis testing. Also illustrate the significance of hypothesis testing as well as sample size. **7**

(b) In the comparison of two kinds of paint, a consumer testing service finds that four 1-gallon cans of one brand cover on the average 546 square feet with a standard deviation of 31 square feet, whereas four 1-gallon cans of another brand cover on the average 492 square feet with a standard deviation of 26 square feet. Assumming that the two population samples are normally distributed and have equal variances, test the null hypothesis $\mu_1 - \mu_2 = 0$ against the alternative hypothesis $\mu_1 - \mu_2 > 0$ at the 0.05 level of significance. **8**

Unit III

5. What do you mean by Power of Visual Perception ? Also discuss the building blocks of Information Visualization. **15**

6. (a) What is meaningful data ? Also write the characteristics of meaningful data. **5**

(b) Write a short note on the history of Information Visualization. **10**

Unit IV

7. Describe the concept of data mining. Also discuss the tools and platforms for data mining. **15**

8. Elaborate the concept of the following clustering techniques : **15**

(a) K-means clustering

(b) Hierarchical clustering.

Unit III

5. What do you mean by Power of Visual Perception ? Also discuss the building blocks of Information Visualization. **15**

6. (a) What is meaningful data ? Also write the characteristics of meaningful data. **5**

(b) Write a short note on the history of Information Visualization. **10**

Unit IV

7. Describe the concept of data mining. Also discuss the tools and platforms for data mining. **15**

8. Elaborate the concept of the following clustering techniques : **15**

(a) K-means clustering

(b) Hierarchical clustering.