Days	No. of Parts Demanded
Mon.	1124
Trues.	1125
Wed.	1110
Thurs.	1120
Fri.	1126
Sat.	1115

Test the hypothesis that the number of parts demanded does not depend on the day of the week. **10**

4. (a) The specifications for a certain kind of ribbon produced by a company, call for a mean breaking strength of 185 pounds. If five pieces randomly selected from different rolls have breaking strengths of 171.6, 191.8, 178.3, 184.9, and 189.1 pounds. Test whether the breaking strength of rolls is less than the company claim at the 0.005 level of significance. **7**

18F115

B.Tech. EXAMINATION, June 2023

(Sixth Semester)

(C-Scheme) (Main & Re-appear)

(Common for all branches)

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, selecting at least *one* question from each Unit. Students can use distribution tables.

Unit I

1. (a) Differentiate between population and sample. Also discuss the different sampling techniques. **7**

(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. **8**

2. (a) Describe the concept of Z-score. Also write its significance. **5**

(b) The median and mode of the following wage distribution are known to be Rs. 3,350 and Rs. 3,400 respectively. Find the values of F3, F4 and F5 : **10**

Wages (in Rs.)	No. of Employees
0-1,000	4
1,000-2,000	16
2,000-3,000	F3
3,000-4,000	F4
4,000-5,000	F4
5,000-6,000	6
6,000-7,000	4
Total	230

Unit II

3. (a) By using the following data regarding pairs (x, y) for variables x and y, find the most likely value of y, when $x = 6.2$ using regression : **5**

(1, 9), (2, 8), (3, 10), (4, 12), (5, 11), (6, 13), (7, 14), (8, 16), (9, 15)

(b) Illustrate the concept of Chi-square test. Given that the demand for a particular spare part in a factory was found to vary from day-to-day. In a sample study the following information was obtained :

(b) Suppose that it is known from experience that the standard deviation of the weight of 8-ounce packages of cookies made by a certain bakery is 0.16 ounce. To check whether its production is under control on a given day, that is, to check whether the true average weight of the packages is 8 ounces., employees select a random sample of 25 packages and find that their mean weight is $\bar{x} = 8.091$ ounces. Since the bakery stands to lose money when $\mu > 8$ and the customer loses out when $\mu < 8$, test the null hypothesis and alternative hypothesis at the 0.01 level of significance. **8**

Unit III

5. (a) What do you understand by data visualization ? Also write its significance. **5**

(b) List and explain the building blocks of information visualization. **10**

6. (a) Illustrate how you can make the abstract data visible. **7**

(b) Briefly discuss the History of information visualization. **8**

Unit IV

7. What is Big data ? Also discuss the concept of In-memory processing along-with its limitations. **15**

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

(a) Agglomerative clustering

(b) DBSCAN clustering.