Total No. of Questions : 4] **SEAT No. :**

PB-103 **[Total No. of Pages : 2**

[6269]-317

T.E. (Computer Engineering) (Insem)

DATA SCIENCE AND BIG DATA ANALYTICS

(2019 Pattern) (Semester - II) (310251)

Time : 1 Hour] ***[Max. Marks : 30***

Instructions to the candidates:

1) ***Answer Q.1 or Q.2, Q.3 or Q.4.***
2) ***Neat diagrams must be drawn wherever necessary.***
3) ***Figures to the right indicate full marks.***
4) ***Assume suitable data if necessary.***
5) ***Use of Scientific Calculator is permitted.***

Q1) a) Explain data wrangling methods with suitable example. **[5]**

b) Suppose that the data for analysis includes the attribute age, given the following data (in increasing order) for the attribute age: 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52,70. **[5]**

i) Use smoothing by bin means, using a bin depth of 3.

ii) What other methods are there for data smoothing?

c) What is data science? Compare data science and information science. **[5]**

OR

Q2) a) Explain 5 V's of Big Data. **[5]**

b) Explain different phases of data analytics life cycle with neat diagram. **[5]**

c) Compare Business Intelligence and data science. **[5]**

Q3) a) Explain skewness and kurtosis. What is the purpose of finding skewness of data? **[5]**

b) What is degree of freedom? Explain with example. **[5]**

c) How hypothesis testing works? Explain steps. **[5]**

OR

P.T.O.

Q4) a) List out measures of dispersion with their significance and mathematical formulae. **[5]**

b) Describe Chi-square Goodness of Fit test. **[5]**

c) Assume that a patient X took a lab test for a certain disease and tested positive. The lab test returns a positive result in 95% of the cases in which the disease is actually present and it falsely returns a positive result in 6% of the cases in which the disease is not present. Further more only 1% of the entire population has this disease. What is the probability that X actually has the disease given that he is tested positive. **[5]**

▽▽▽▽