USN | | | | | | | | | |

Sixth Semester B.E. Degree Examination, July/August 2022
Data Analytics for Engineers

Time: 3 hrs.   Max. Marks: 100

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. Use of statistical tables is permitted.

Module-1

1 a. Why R language required in statistical data? **(05 Marks)**
   b. List the following points describe why R language should be used. **(10 Marks)**
   c. List the advantages and disadvantages of R language. **(05 Marks)**

OR

2 a. Explain the following terms with respect to R:
   i) Factor   ii) Data frame   iii) Vector   iv) Variable   v) Function. **(05 Marks)**
   b. Explain briefly commonly used method of data input, with an example. **(10 Marks)**
   c. Explain briefly R statistical software and mention what are the packages available. **(05 Marks)**

Module-2

3 a. Explain the following terms:
   i) Mean   ii) Median   iii) Mode   iv) Standard deviation   v) Stern and leaf diagram. **(05 Marks)**
   b. The following table shows the frequency distribution of college students according to their pocket money (daily). Explain R commands for hixtogram and frequency polygon.

Pocket-Money (in Rs.)	20-29	30-39	40-49	50-59	60-69	70-79	80-89
Number of students	10	24	18	12	8	5	3

(10 Marks)
   c. List the properties of standard deviation. **(05 Marks)**

OR

4 a. Explain the following terms with sketches:
   i)   Symmetrical skewed
   ii) Positively skewed
   iii) Negatively skewed
   iv) Co-efficient of skewness
   v)   Bowley's measures of skewness. **(10 Marks)**
   b. The following tables shows the number of hours (Y) slept by 45 hospital patients following the administration of a certain drug. Using R command explain person's measures of skewness [No out put]. **(10 Marks)**

y	1	2	3	4	5	7	8	10	11	12	13	17
f	4	2	7	5	3	8	6	3	2	2	2	1

Module-3

5 a. State the laws of probability. **(06 Marks)**

b. Explain binomial distribution with its probability mass function and mention its properties.
(06 Marks)

c. Explain Poisson distribution with an example. List the uses. **(08 Marks)**

OR

6 a. Explain R commands used for normal distribution. **(08 Marks)**

b. Write the R commands to determine the probability for the given problem:
i) $P(x \leq 2)$ ii) $P(0.84 \leq x \leq 2.5)$ iii) $P(x \geq 2)$ [No output]. **(06 Marks)**

c. List the properties of normal distribution curve. Mention its importance. **(06 Marks)**

Module-4

7 a. Explain R-function for t-distribution. **(08 Marks)**

b. State central limit theorem. **(02 Marks)**

c. In a lake pollution study the concentration of lead in the upper sedimentary layer of a lake bottom is measured from 36 sediment samples of 1000 cubic centimeters each. The sample mean and standard deviation are found to be 0.38 and 0.06 respectively. Test the hypothesis that the mean concentration of lead in the upper sedimentary layer of the lake bottom is 0.34 per 1000 cubic centimeters. **(10 Marks)**

OR

8 a. List the general procedure for hypothesis test. **(08 Marks)**

b. Explain:
i)   Type – I error
ii)   Type – II error
iii)   Level of significance. **(06 Marks)**

c. Explain concept of P-value. **(06 Marks)**

Module-5

9 a. Define and explain different types of correction. **(10 Marks)**

b. List the properties and interpretation of correlation of co-efficient. **(05 Marks)**

c. Explain with R commands for scatter diagram and compute the correlation co-efficient between amount (x) of fertilizer and the yield (y) of potatoes for the given data. [No output].

Amount x :	0	4	8	12
Yield y :	8.34	8.89	9.16	9.50

(05 Marks)

OR

10 a. Define and mention properties of regression analysis. **(05 Marks)**

b. Define scatter diagram and explain linear regression model. **(08 Marks)**

c. Explain polynomial regression models. **(07 Marks)**

* * * * *