

U.G. 6th Semester Examination - 2022

COMPUTER SCIENCE

[HONOURS]

Discipline Specific Elective (DSE)

Course Code : COM.SC-H-DSE-L-603

Full Marks : 60

Time : $2\frac{1}{2}$ Hours

The figures in the right-hand margin indicate marks.

Candidates are required to give their answers in their own words as far as practicable.

GROUP - A

1. Answer any **ten** questions: 2×10=20
- a) What is contingency table?
 - b) Explain the role of Git in software development.
 - c) What are the ways of installing a package in R?
 - d) Write a code to create the matrix $\begin{bmatrix} 9 & 3 \\ 7 & 5 \end{bmatrix}$ using R.
 - e) Why is standard deviation preferred over variance?
 - f) Explain the role of sampling in data science.
 - g) What do we measure with skewness?
 - h) Why is data science called multidisciplinary?

[Turn Over]

- i) How to calculate the median of a collection of numbers?
- j) Briefly explain box plot.
- k) "Outliers affect mean but not median"– Explain.
- l) What are the different modes of opening a file in R?
- m) Create the following section and subsection using RMarkdown:

1. Section One

1.1. Subsection One

1.1.1 Sub Subsection One

1.2 Subsection Two

- n) Create a sample unordered list using RMarkdown.
- o) What is reproducibility crisis?

GROUP-B

Answer any **four** questions : 5×4=20

- 2. Briefly explain the data science process with a suitable diagram. 5
- 3. Why should research be reproducible? Justify your answer. 5
- 4. What is histogram plot of data? How can different measures of central tendencies be depicted on histogram? 2+3

5. Which functions in R are used to convert strings of date format into date type data?

Consider on 31st July, 2022 at 10 PM IST you will be catching a flight from Kolkata and you will reach New York on 1st August, 2022 at 8:30 AM ET. Use R to store the above two dates and find the time taken in transit. 2+3

6. Consider the set of numbers– 32; 45; 19; 78; 24. Find mean, median and standard deviation of the set.

1+1+3

7. Explain the role of knitr in reproducible research with R. 5

GROUP-C

Answer any **two** questions : 10×2=20

8. Explain the importance of data cleaning by discussing a few techniques of it. 10

9. Briefly explain a few techniques to visualize one and two- dimensional data. 5+5

10. Write your own recursive function in R that will print the sum of n numbers. The number n will be passed to the function as an argument. 10

11. a) What do you mean by dimensions of data?
b) What is dimensionality reduction of data?

- c) Why reduction of dimensions is important in data science?

- d) Briefly explain one technique of dimensionality reduction. 2+2+2+4
