



VI SEMESTER B.TECH.

END SEMESTER EXAMINATIONS, APRIL 2018

SUBJECT: PROGRAM ELECTIVE: DATA SCIENTIST'S TOOLBOX
AND R PROGRAMMING

[CRA 4009]

REVISED CREDIT SYSTEM
(24/04/2018)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

- ❖ Answer ALL the questions.
- ❖ Missing data if any, may be suitably assumed.

- 1A. What are the indications to say something is wrong in R? What are the primary tools for debugging in R? Explain. 5
- 1B. When do we use mapply function? What is the output of mapply(rep,1:4,4:1)? 3
- 1C. Will the following code snippet run? If yes, what is the output? If not what is the error/warning? 2
- ```

x <- 1:10
if(x > 5) {
 x <- 0
}

```
- 2A. Simulate a linear model for 100 observations of  $x$  and  $\varepsilon$  such that  $y = \beta_0 + \beta_1 x + \varepsilon$  where  $N(0,2^2)$ ,  $x \in N(0,1^2)$ ,  $\beta_0 = 0.5$  and  $\beta_1 = 2$  5
- 2B. Explain with example any three prediction key quantities. 3
- 2C. Write R code to download a dataset from any website, skip first 15 rows and display next 10 rows of data from the dataset. 2
- 3A. Explain the following functions with an example: 5
- i. strsplit      iv. format
  - ii. sapply      v. grep
  - iii. gsub
- 3B. Explain different types of data with an example for each. Write the characteristics of tidy data. 3

- 3C. Considering the following code, how much amount of time in percentage is spent in the `lm` function, based on the `by.total` method of normalization in summary `Rprof()`? Justify your answer? 2
- ```
library(datasets)
Rprof()
fit <- lm(y ~ x1 + x2)
Rprof(NULL)
```
- 4A. What is meant by web scraping? Discuss any two ways through which web can be scraped using R. Explain with example how to access websites with password. 5
- 4B. Consider a data frame with fields Name, Age, Salary, No_Yrs_Exp, BankLoan 3
- Write R code to perform the following:
- Display age of persons with top 10 salaries.
 - Display names of persons whose salary is more than Rs. 10,000.
 - Add a new field department to the data frame and display only those rows where department is either 'CS' or 'ECE'.
- 4C. What are free variables? Identify free variable/s from the following code snippet 2
- ```
h <- function(x, y = NULL, d = 3L) {
 z <- cbind(x, d)
 if(!is.null(y))
 z <- z + y
 else
 z <- z + f
 g <- x + y / z
 if(d == 3L)
 return(g)
 g <- g + 10
 g
}
```
- 5A. List and explain different types of data science questions. 5
- 5B. Write code snippet in R to perform the following: 3
- To match "MIT" at the start and end of a line.
  - To match a list of alphabets.
  - To match a sentence which starts with a digit followed by characters or digits.
- 5C. Write R code to connect to a database and count the number of tuples in any table of the database. 2