

INTRODUCTION TO R
(CSEN 2204)

Time Allotted : 2½ hrs

Full Marks : 60

Figures out of the right margin indicate full marks.

*Candidates are required to answer Group A and
any 4 (four) from Group B to E, taking one from each group.*

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

Group - A

1. Answer any twelve:

12 × 1 = 12

Choose the correct alternative for the following

- (i) What will be displayed, when the statement `as.logical(1)` is executed in R?
(a) TRUE (b) FALSE (c) Numeric 1 (d) ASCII code of 1.
- (ii) Which of the following can be considered as object attribute in R?
(a) dimension (b) class (c) length (d) All of (a), (b) & (c).
- (iii) What will be the output of the following R code?
`x <- c(1, NA, 4, NA, 1)`
`y <- c(NA, 5, 1, NA, -3)`
`x - y`
 (a) 1 -5 5 0 4 (b) (1-NA) (NA-5) 3 0 4
 (c) NA NA 3 NA 4 (d) NA NA 3 0 4
- (iv) What will be displayed when the following piece of code is executed in R?
`c(13, 23, 33) %in% c(3, 13, 23, 53, 73)`
 (a) FALSE TRUE TRUE FALSE FALSE (b) `c(13, 23)`
 (c) `c(3, 13, 23, 33, 53, 73)` (d) TRUE TRUE FALSE
- (v) Which of the following is a valid variable name in R?
(a) MUX2:1 (b) MUX2-1 (c) 2_1_MUX (d) MUX2_1.
- (vi) What will be displayed when the following code snippet is run in R?
`v1 <- c(2L, 'xyz', TRUE, 5.7)`
`typeof(v1)`
 (a) "integer" (b) "character" (c) "logical" (d) "double"
- (vii) Point out the correct statement with respect to `dbSendQuery()`
 (a) The query is limited to INSERT and DELETE statements only
 (b) It sends any SQL query and does not return any result object
 (c) The query is limited to UPDATE statement only
 (d) The query is limited to SELECT statement only
- (viii) Which of the following cannot be the output when the statement `sample(1:6, size = 2)` is executed in R?
 (a) 1 6 (b) 6 1 (c) 3 3 (d) 2 5
- (ix) What will be the output of the following R Code?
`x <- 1:4`
`x > 2`
 (a) 3 4 (b) 1 2 3 4
 (c) FALSE FALSE TRUE TRUE (d) TRUE TRUE
- (x) When an object is called in the command line, R will look for the object, first in the
 (a) current active environment (b) parent of the current active environment
 (c) child of the current active environment (d) global environment.

Fill in the blanks with the correct word

- (xi) In R, missing values in a data frame are represented by _____.
- (xii) The output of the statement `is.na(c(1, 2, NA, 3, NA, 4))` is _____.
- (xiii) The output of `sqrt(-5)` in R is _____.
- (xiv) The statement `raw(3)`, when executed in R, will display _____.
- (xv) The statement `rep(x = 4:6, length.out = 8)` when executed in R, will display _____.

Group - B

2. (a) Explain what will be displayed when the following code snippet is run in R.

```
vec<- c(0, 0, 0, 0, 0, 0)
vec[c(2, 4, 6)] <- c(1, 3, 5)
vec[3:5] <- vec[3:5] + 2
print(vec)
```

[[CO2](Apply/IOCQ)]
- (b) Explain what will be displayed when the following statements are executed in R.

```
matrix(1:8, ncol = 4)
class("HITK")
df<- data.frame(Student = c("Priya", "Sajal", "Abhi"),
                Physics = c(89, 91, 94), Chemistry = c(95, 84, 90),
                Mathematics = c(90, 99, 96))
df[2, c(TRUE, TRUE, FALSE, TRUE)]
```

[[CO2](Apply/IOCQ)]
- (c) What is an R Script? What is its advantage? [[CO1](Understand/LOCQ)]
- (d) Explain with examples, four different types of vectors in R programming language? [[CO1](Remember/LOCQ)]
- 2 + (1 + 1 + 2) + 2 + 4 = 12**

3. (a) Write a function 'prime_numbers' in R to get all prime number upto a given number. Call this function to display the list of prime numbers upto 25. [[CO3](Analyse/HOCQ)]
- (b) There is an atomic vector c("I", "study", "in", "HITK"). Write an R code, using for-loop, to display each of the items in separate lines and also display all the items in one line, separated by spaces. [[CO1](Apply/IOCQ)]
- (c) Write a function in R which takes an integer vector and returns the mean of the elements of the vector, without using any in-built function. Use 'repeat' loop in the code. Call this function to display the mean of the numbers 1, 5, 9, 15, 27. [[CO3](Apply/IOCQ)]
- (3 + 1) + (2 + 2) + (3 + 1) = 12**

Group - C

4. (a) Explain what will be displayed when the following code snippet is run in R.

```
data = c(7: 11, 54, 43, 54, 89, 50)
print(data)
print(summary(data))
```

[[CO4](Analyse/HOCQ)]
- (b) What are the arguments of the following functions in R? What do they return?
 (i) rnorm() (ii) pnorm() [[CO4](Remember/LOCQ)]
- (c) Explain with suitable example why \$ notation is used in R. What is the difference between \$ notation and double brackets [[]]? [[CO2](Remember/LOCQ)]
- (1 + 3) + (2 + 2) + (2 + 2) = 12**

5. (a) Explain with an example what is 'factor' in R. [[CO2](Understand/LOCQ)]
- (b) Consider the data of customers of a Fast Food shop given in the following table:

Age (yrs)	Weight (Kg)	Annual Income (Rs in Lakh)	Gender
21	52	2.52	Male
33	60	7.20	Female
28	52	4.20	Female
39	72	8.64	Male
34	80	6.36	Male

Write an R code to

- (i) Create a data frame with name 'customers'
- (ii) Convert the gender of the customers into factors and convert them into numeric values
- (iii) Obtain sorted unique values of the column 'Weight'
- (iv) Extract and display only the Annual Income of the customers. [[CO2](Analyse/HOCQ)]
- (c) Write a statement in R, which will create a list of three elements containing
- (i) a character vector "HITK" as the first element,
- (ii) a numeric vector containing a sequence of 10 numbers from 5 to 14 as the second element, and
- (iii) a list of 2 logical objects TRUE and FALSE, as the third element. [[CO2](Understand/LOCQ)]
- 2 + (4 × 2) + 2 = 12**

Group - D

6. (a) What is the difference between dbExecute() and dbSendStatement() in DBI package? [[CO5](Understand/LOCQ)]
- (b) Explain the roles of the following commands in DBI package, with an example in each case
 (i) dbConnect (ii) dbBind (iii) dbFetch [[CO5](Remember/LOCQ)]
- (c) Write the dplyr commands and corresponding SQL statements for the following queries:
- (i) counting the total number of rows in the dataset
- (ii) returning the top n rows of the dataset. [[CO4](Apply/IOCQ)]
- 2 + (3 × 2) + (2 + 2) = 12**

7. (a) What does the function `md.pattern()` do in mice package? Explain with an example. [[CO4](Understand/LOCQ)]
 (b) What are the different types of missing data? Why is it required to handle them? [[CO4](Remember/LOCQ)]
 (c) Describe briefly with examples, how imputation is done with R missForest Package. [[CO4](Apply/IOCQ)]
- 3 + (2 + 2) + 5 = 12**

Group - E

8. (a) The 'computer' dataset contains 200 rows and 5 columns. The columns indicate Manufacturer, Model, specInt, specFP and Price.
 Write an R code, using ggplot2 package, to draw a scatter plot with specInt variable on the x-axis and the specFP variable on the y axis, with different colours for different manufacturers. Explain your code. [[CO6](Analyse/HOCQ)]
- (b) (i) Write an R code to do the following:
 Create a vector with all the alphabets from the word "mississippi"
 Generate a simple frequency table
 Visualize the above frequency table in graphics (using ggplot2)
 Generate a frequency table with Proportions
 Generate a cumulative frequency table
- (ii) What would be the output of the cumulative frequency table? Sketch the visualization of the simple frequency table in graphics and of the? [[CO6](Apply/IOCQ)]
- 4 + (5 + 3) = 12**
9. (a) Explain, with suitable examples, how the melt() and cast() functions are used in R to efficiently reshape the data. [[CO6](Understand/IOCQ)]
 (b) Describe briefly 4 different types of data visualization in Data Science. [[CO6](Remember/LOCQ)]
 (c) How density curve is drawn in ggplot2? [[CO2](Apply/IOCQ)]
- 4 + 6 + 2 = 12**

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	40.63	38.54	20.83

Course Outcome (CO):

After completion of the course, students will be able to:

- CSEN2204.1. Learn and understand the basics of the R Programming Language.
- CSEN2204.2. Learn about basic R data structures.
- CSEN2204.3. Learn about how to develop reusable modules in R and apply them.
- CSEN2204.4. Use various libraries and packages of R Programming.
- CSEN2204.5. Learn about data exploration, querying in R.
- CSEN2204.6. Learn how to visualize data and use graphics in R.

*LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.

