INTRODUCTION TO R
(CSEN 2204)

Time Allotted : 3 hrs **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.

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

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

(i) In R, variables are
(a) Statically typed
(b) Dynamically typed
(c) Whether statically or dynamically typed depends on the variable
(d) Whether statically or dynamically typed depends on the R version being used.

(ii) For a frequency distribution of a variable p, mean = 45, median= 40. The distribution is
(a) Positively skewed (b) Negatively skewed
(c) Symmetrical (d) None of these.

(iii) What will be the output of the following R code?
sqrt(-45)
(a) 6.708 (b) -6.708
(c) NaN (d) None of these.

(iv) R does not have a standard in-built function to create which of the following?
(a) Mean (b) Median (c) Mode (d) All of (a), (b) & (c).

(v) In multiple linear regression, we have
(a) Multiple predictor variables and multiple response variables
(b) Multiple predictor variables and single response variable
(c) Single predictor variable and multiple response variables
(d) Whether there will be multiple predictor and response variables depend on the data set being analysed.

(vi) In R, missing values in a data frame are represented by
(a) "NA" (b) "$"
(c) "_" (d) " " (Blank space).

(vii) In a boxplot, the horizonal line running inside the box indicates the
(a) Mean (b) Median (c) Mode (d) Standard Deviation.

(viii) What will be the output of the following R code?
> vector <- c(3, 5, 1, 6, 12, 4)
> which(vector > 5)
(a) 6 12 (b) 5 6 (c) 4 5 (d) 3 4.

(ix) What will be the output of the following R code?
x <- c(4, 7, NA, 4, 8,1)
y <- c(5, NA, 1, 2, 3,1)
x + y
(a) 9 7 1 6 11 2 (b) 9 NA NA 6 11 2
(c) 24 (d) Error.

(x) Consider the following matrix m2.
m2 = matrix(1:10, nrow = 5)
What will be the output of the following R code?
a = m2[1, 2:3]
(a) subscript out of bounds (b) 1 6
(c) 3 8 (d) None of (a), (b) & (c).

Group - B

2. (a) Is R an interpreted language or a compiled language? What are the specific advantages and disadvantages of the way R code is executed? *[(CO1)(Understand/LOCQ)]*

(b) Which of the following variable names are not valid in R, and why?
(i) _my_variable
(ii) my_variable
(iii) my_variale?
(iv) 5my_variable. *[(CO1)(Understand/LOCQ)]*

(c) What are the outputs of the following code snippets?
(i)
```
> fruits <- c("apple","orange","guava")
> print(class(fruits))
```
(ii)
```
> A <- matrix(c(5:16), nrow = 4,ncol=3)
> B <- matrix(c(1:12), nrow = 4,ncol=3)
> sum <- A+B
> print(sum).
```
[(CO2) (Evaluate/HOCQ)]

(1 + 3) + (1 + 1 + 1 + 1) + (2 + 2) = 12

3. (a) Write an R program to create a data frame of Employees using the following data:

Age	Height	Weight	Gender
23	76	50	Female
21	62	52	Female
34	63	80	Male
44	69	65	Male
32	72	70	Female

Write R code to
(i) Extract and display only the weight of the Employees data frame.
(ii) Convert the Gender of the Employees into factors and convert them into numeric values.
(iii) Obtain unique values of the column Age.
(iv) Obtain the sorted unique values of the column Age.
(v) Delete the Height column. *[(CO1) (Remember/LOCQ)]*

(b) Write an R program to create a sequence of numbers from 20 to 50 and find the mean of numbers from 20 to 60 and sum of numbers from 51 to 91. *[(CO1)(Understand/IOCQ)]*

(2 + 2 + 2 + 1 + 2 + 1) + 2 = 12

Group - C

4. (a) What is a factor in R? Explain with examples. What are the steps in creating a factor? *[(CO2) (Remember/LOCQ)]*

(b) What is a data frame in R? What are a data frame's components? What kind of information about the data frame can be obtained by applying the summary() function? *[(CO2) (Analyze/IOCQ)]*

(c) (i) How can you load a image file in R?
(ii) In R, how can you have a function return multiple objects as output? Explain with an example. *[(CO3) (Analyze/IOCQ)]*

(2 + 1 + 1) + (2 + 1 + 1) + (2 + 2) = 12

5. (a) What is lazy evaluation of a function in R? Explain with examples. What is an inline function? *[(CO3) (Understand/LOCQ)]*

(b) How to create a Histogram? Write a R program to get all prime numbers up to a given number. *[(CO3) (Understand/LOCQ)]*

(c) (i) Given a dependent variable (Y) and an independent variable (X), how does linear regression establish a relation between X and Y through an equation? Explain all relevant terms in the relationship.
(ii) What is covariance? How does it differ from variance? *[(CO5) (Understand/LOCQ)]*

(3 + 1) + (2 + 2) + (2 + 2) = 12

Group - D

6. (a) You are given the following code:
```
x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152,131)
y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
relation <- lm(y~x)
print(summary(relation))
```
And this is the output from running the above code:
Call:

```
lm(formula = y ~ x)
Residuals:
Min 1Q Median 3Q Max
-6.3002 -1.6629 0.0412 1.8944 3.9775
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -38.45509 8.04901 -4.778 0.00139 **
x 0.67461 0.05191 12.997 1.16e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 3.253 on 8 degrees of freedom
Multiple R-squared: 0.9548, Adjusted R-squared: 0.9491
F-statistic: 168.9 on 1 and 8 DF, p-value: 1.164e-06
```

Explain in detail, the meaning of the last line of the output "F-statistic: 168.9 on 1 and 8 DF, p-value: 1.164e-06". *[(CO6) (Evaluate/HOCQ)]*

(b) What is skewness and why is it important? For a distribution that is negatively skewed, would the mean be higher than the median, or vice versa? Explain your answer. *[(CO5) (Understand/LOCQ)]*

(c) What is the output of the in-built R function rnorm()? What are the parameters of this function? What does the pnorm() function generate? What is pnorm(78, mean = 74, sd = 2, lower.tail = FALSE) supposed to specifically calculate? *[(CO4) (Understand/LOCQ)]*

4 + (2 + 2) + (1 + 1 + 1 + 1) = 12

7. (a) What is data reshaping in R? Why is it important? What are the different methods of data reshaping? *[(CO2) (Remember/LOCQ)]*

(b) What is AIC? How is it applied? Which package and function of R can be used to calculate AIC? *[(CO5) (Understand/LOCQ)]*

(c) Sketch the output when the following code is executed. What is the name of the plot that the code generates?

```
> x <- c(5,7,8,7,2,2,9,4,11,12,9,6)
> y <- c(99,86,87,88,111,103,87,94,78,77,85,86)
> plot(x, y).
```

[(CO5) (Analyze/IOCQ)]

(1 + 2 + 1) + (1 + 1 + 2) + (3 + 1) = 12

Group - E

8. (a) (i) How are missing values usually represented in R?
(ii) Given a data frame called my_data, write a piece of R code to remove empty rows and columns from data. *[(CO2) (Apply/IOCQ)]*

(b) Which type of a plot can help detect outliers in a dataset? Explain with examples. What is the function in R for generating such a plot? *[(CO5)(Understand/LOCQ)]*

(c) What are raster images? How are they different from vector images? *[(CO5)(Remember/LOCQ)]*

(1 + 3) + (1 + 2 + 1) + (2 + 2) = 12

9. (a) Write a function in R to take as input an integer *n* and return the value of the *n*'th Fibonacci number as the output using the following approaches:
(i) Iteration
(ii) Recursion. *[(CO4) (Analyze/IOCQ)]*

(b) You are given the runs scored in each test match innings batted by Sir Donald Bradman and Sachin Tendulkar. Using this data and all the R in-built functions and programming techniques you are aware of, design a study that will help you establish whether Bradman or Tendulkar was a better batsman. Write down the steps of the study with supporting R code. *[(CO5)(Create/HOCQ)]*

(4 + 4) + 4 = 12

Cognition Level	*LOCQ*	*IOCQ*	*HOCQ*
Percentage distribution	*60.42*	*27.08*	*12.5*

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*