**SAVEETHA SCHOOL OF ENGINEERING**

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES**

**ITA 0443 - STATISTICS WITH R PROGRAMMING FOR REAL TIME PROBLEM**

**DAY 2 – LAB ASSESSMENT**

**Reg No:192124095**

**Name:K.Abdul Masood**

1.Write a R program to create an array of two 3x3 matrices each with 3 rows and 3 columns from two given two vectors. Print the second row of the second matrix of the array and the element in the 3rd row and 3rd column of the 1stmatrix.   
Program:

a <- c(1, 2, 3, 4, 5, 6, 7, 8, 9)

b <- c(10, 20, 30, 40, 50, 60, 70, 80, 90)

mat1 <- matrix(a, nrow = 3, ncol = 3, byrow = TRUE)

mat2 <- matrix(b, nrow = 3, ncol = 3, byrow = TRUE)

mat\_array <- array(c(mat1, mat2), dim = c(3, 3, 2))

print(mat\_array[2, , 2])

print(mat\_array[3, 3, 1])

Output:

[1] 40 50 60

[1] 9

**2.** Write a R program to combine three arrays so that the first row of the first array is followed by the first row of the second array and then first row of the third array.

Program:

array1 <- array(1:3, dim = c(1, 3))

array2 <- array(4:6, dim = c(1, 3))

array3 <- array(7:9, dim = 3)

combined\_array <- rbind(array1, array2, array3)

print(combined\_array)

Output:

[1] [2] [3]

1 2 3

4 5 6

7 8 9

**3.** Write a R program to create an array using four given columns, three given rows, and two given tables and display the content of the array.   
Program:

table1 <- data.frame(col1 = c(1, 2, 3), col2 = c(4, 5, 6))

table2 <- data.frame(col3 = c(7, 8, 9), col4 = c(10, 11, 12))

array\_data <- array(c(table1$col1, table1$col2, table2$col3, table2$col4), dim = c(3, 4))

print(array\_data)

Output:

[1] [2] [3] [4]

[1] 1 4 7 10

[2] 2 5 8 11

[3] 3 6 9 12

**4.** Write a R program to create a two-dimensional 5x3 array of sequence of even integers greater than50. 

**Use Below Data frame from question 5 to 9**

exam\_data = data.frame(

name = c('Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'),

score = c(12.5, 9, 16.5, 12, 9, 20, 14.5, 13.5, 8, 19),

attempts = c(1, 3, 2, 3, 2, 3, 1, 1, 2, 1),

qualify = c('yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes')

)

Program

exam\_data data frame

subset\_data <- exam\_data[c(3, 5), c(1, 3)]

print(subset\_data)

5. Write a R program to extract 3rd and 5th rows with 1st and 3rd columns from a given data frame.

Program:

exam\_data data frame

subset\_data <- exam\_data[c(3, 5), c(1, 3)]

print(subset\_data)

6. Write a R program to add a new column named country in a given data frame

Country<-c("USA","USA","USA","USA","UK","USA","USA","India","USA","USA")

Program:

country <- c("USA", "USA", "USA", "USA", "UK", "USA", "USA", "India", "USA", "USA")

exam\_data$country <- country

print(exam\_data)

7. Write a R program to add new row(s) to an existing data frame

new\_exam\_data = data.frame(name = c('Robert', 'Sophia'),score = c(10.5, 9), attempts = c(1, 3),qualify = c('yes', 'no'))

Program:

exam\_data <- rbind(exam\_data, new\_exam\_data)

print(exam\_data)

8. Write a R program to sort a given data frame by name and score.

Program:

exam\_data <- exam\_data[order(exam\_data$name),]

exam\_data <- exam\_data[order(exam\_data$score),]

print(exam\_data)

9.  Write a R program to save the information of a data frame in a file and display the information of the file.

Program:

write.csv(exam\_data, file = "exam\_data.csv")

saved\_exam\_data <- read.csv("exam\_data.csv")

print(saved\_exam\_data)

10. Write a R program to call the (built-in) dataset airquality. Check whether it is a data frame or not? Order the entire data frame by the first and second column. remove the variables 'Solar.R' and 'Wind' and display the data frame.

Program:

data("airquality")

is.data.frame(airquality)

airquality <- airquality[order(airquality[,1], airquality[,2]),]

airquality <- airquality[, !(names(airquality) %in% c("Solar.R", "Wind"))]

print(airquality)

Output:

|  |
| --- |
| Ozone Temp Month Day  21 1 59 5 21  23 4 61 5 23  18 6 57 5 18  76 7 80 7 15  147 7 69 9 24  11 7 74 5 11  9 8 61 5 9  94 9 81 8 2  137 9 71 9 14  114 9 72 8 22  73 10 73 7 12  20 11 62 5 20  13 11 66 5 13  22 11 73 5 22  50 12 73 6 19  3 12 74 5 3  141 13 76 9 18  138 13 71 9 15  51 13 76 6 20  144 13 64 9 21  148 14 63 9 25  151 14 75 9 28  14 14 68 5 14  16 14 64 5 16  82 16 74 7 21  95 16 82 8 3  143 16 82 9 20  12 16 69 5 12  15 18 58 5 15  152 18 76 9 29  140 18 67 9 17  4 18 62 5 4  8 19 59 5 8  49 20 65 6 18  87 20 82 7 26  153 20 68 9 30  130 20 80 9 7  47 21 77 6 16  132 21 75 9 9  113 21 77 8 21  135 21 76 9 12  108 22 77 8 16  28 23 67 5 28  145 23 71 9 22  110 23 76 8 18  44 23 82 6 13  131 23 78 9 8  7 23 65 5 7  142 24 68 9 19  133 24 73 9 10  74 27 81 7 13  136 28 77 9 13  105 28 82 8 13  6 28 66 5 6  38 29 82 6 7  149 30 70 9 26  19 30 68 5 19  111 31 78 8 19  24 32 61 5 24  129 32 84 9 6  64 32 81 7 3  17 34 66 5 17  78 35 82 7 17  97 35 85 8 5  2 36 72 5 2  146 36 81 9 23  31 37 76 5 31  48 37 72 6 17  93 39 81 8 1  41 39 87 6 10  67 40 83 7 6  1 41 67 5 1  112 44 78 8 20  104 44 86 8 12  134 44 81 9 11  116 45 79 8 24  29 45 81 5 29  139 46 78 9 16  128 47 87 9 5  77 48 81 7 16  63 49 85 7 2  90 50 86 7 29  88 52 86 7 27  109 59 79 8 17  92 59 81 7 31  79 61 84 7 18  81 63 85 7 20  66 64 83 7 5  91 64 83 7 30  106 65 80 8 14  98 66 87 8 6  40 71 90 6 9  126 73 93 9 3  118 73 86 8 26  120 76 97 8 28  68 77 88 7 7  125 78 92 9 2  96 78 86 8 4  80 79 87 7 19  85 80 86 7 24  89 82 88 7 28  122 84 96 8 30  71 85 89 7 10  123 85 94 8 31  100 89 90 8 8  127 91 93 9 4  124 96 91 9 1  69 97 92 7 8  70 97 92 7 9  86 108 85 7 25  101 110 90 8 9  30 115 79 5 30  121 118 94 8 29  99 122 89 8 7  62 135 84 7 1  117 168 81 8 25  60 NA 77 6 29  58 NA 73 6 27  53 NA 76 6 22  107 NA 79 8 15  25 NA 57 5 25  54 NA 76 6 23  59 NA 80 6 28  65 NA 84 7 4  57 NA 78 6 26  56 NA 75 6 25  103 NA 86 8 11  61 NA 83 6 30  72 NA 82 7 11  150 NA 77 9 27  52 NA 77 6 21  119 NA 88 8 27  35 NA 84 6 4  10 NA 69 5 10  36 NA 85 6 5  102 NA 92 8 10  34 NA 67 6 3  43 NA 92 6 12  55 NA 76 6 24  115 NA 75 8 23  83 NA 81 7 22  42 NA 93 6 11  37 NA 79 6 6  26 NA 58 5 26  39 NA 87 6 8  32 NA 78 6 1  33 NA 74 6 2  75 NA 91 7 14  84 NA 82 7 23  46 NA 79 6 15  45 NA 80 6 14  5 NA 56 5 5  27 NA 57 5 27 |
|  |
| |  | | --- | |  | |

11. Write a R program to create a factor corresponding to height of women data set , which inbuild in R, contains height and weights for a sample of women.

Program:

data("women")

height <- women[,2]

height\_factor <- factor(height)

print(levels(height\_factor))

Output:

[1] "115" "117" "120" "123" "126" "129" "132" "135" "139" "142" "146"

[12] "150" "154" "159" "164"

12. Write a R program to extract the five of the levels of factor created from a random sample from the LETTERS (Part of the base R distribution.)

Program:

letters\_factor <- sample(LETTERS, 5)

letters\_factor <- factor(letters\_factor)

levels(letters\_factor)

print("levels")

Output:

[1] "levels”

13. **Iris** dataset is a very famous dataset in almost all data mining, machine learning courses, and it has been an R build-in dataset. The dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features(variables) were measured from each sample, they are the **length** and the **width** of sepal and petal, in centimetres. Perform the following EDA steps .

(i)Find dimension, Structure, Summary statistics, Standard Deviation of all features.

(ii)Find mean and standard deviation of features groped by three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor)

(iii)Find quantile value of sepal width and length

(iV)create new data frame named iris1 which have a new column name **Sepal.Length.Cate** that categorizes “Sepal.Length” by quantile

(V) Average value of numerical varialbes by two categorical variables: Species and Sepal.Length.Cate:

(vi) Average mean value of numerical varialbes by Species and Sepal.Length.Cate

(vii)Create Pivot Table based on Species and Sepal.Length.Cate.

14.Titanic Casualties – Use the standard ‘Titanic’ dataset which is part of R Base to answer  
the following questions.  
(i). Use an appropriate apply function to get the sum of males vs females aboard.  
(ii). Get a table with the sum of survivors vs sex.  
(iii). Get a table with the sum of passengers by sex vs age