Expt. No. _____ Date _____

Page No. _____

Experiment - 1

Aim:

Introduction to Basic R-programming and familiarization with the concept

Programs (Basic Declaration)

1.
```
name = readline(prompt = "Input your Name")
age = readline(prompt = "Input your age")
print(paste("My name is ", name, "and I am", age,
            "years old"))
print(R.version.string)
```

2.
```
a = 4
b = 6
c = a+b
print(c)
```

3.
```
n = readline(prompt("Input a number."))
x = 0
for(i in 1:n) {
    if(i %%. %1.3 == 0) {
        x = x+1
    }
}
```

Teacher's Signature _____

4. m = readline(prompt = "Input a number")

```
x=0
for (i in 1:m){
    if ((i%.%3==0 && i%.%5==0) ||
        (i%.%3==0 && i%.%7==0)) {
    x=x+1
    }
}
```

5. Squares of numbers in sequence.

```
n=readline(prompt="Enter a number")
for(i in 1:n) {
    print(i^2)
}
```

6. Create a list of 5 numbers (10, 30, 50, 80, 110)
Find mean, median

```
x = c(10,30,50,90,110)
mean(x)
median(x)
```

Teacher's Signature _____

Output:

1) Input your name: Manasa
Input your age: 21
My name is Manasa and I am 21 years old.

2) 10

3) Input a number: 30
10

4) Input a number: 50
5

5) Enter a number: 5
1
4
9
16
25

6) 58
50

Expt. No. _____ Page No. _____

Experiment - 2.

Aim:

Write a program in R for the following:

1) R program to find read "testfile.csv"
2) Store values in array
3) Provide lineplot B against A
4) Compute D = (colB + col C) | (col A)
5) Provide line plot D against A
6) Compute E = average A, B, C, D
7) Provide list of rows when value of E >
8) Sort in ascending order
9) Write newly added columns to csv file.
10) Create new file with subset of rows from 7 to 8

Code:

```
df = read.csv("\\Users\\Manasa\\Downloads\\testfile.csv")
lst1 = list()
for (i in 1:ncol(df)){
        list[[i]] <- df[i,]
}
arr = array(unlist(list1), dim=(15,3))
print(arr)

arr = array(unlist(list1), dim = c(15,5))
  for(i in 1:15){
```

Teacher's Signature _____

Expt. No. _____

```
arr[i,5] <-(arr[i,1] + arr[i,2] + arr[i,3] +
            arr[i,4])/4
}
print(arr)
 for(i in 1:15){}
plot(arr[,4], arr[i], type=1) med <- median(arr[5])
print(med)
 for(i in 1:15){
        if(arr[i,5] >med){
            print(arr[i,3])
    }
  }

sort(arr[,5], decrease = FALSE)
write csv(arr, "Users\Manav\Downloads\testfile.csv")
print(arr)
```

Teacher's Signature _____

Output:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	1	1	1	2	1.25
[2,]	2	4	8	6	5.00
[3,]	3	9	27	12	12.75
[4,]	4	16	64	20	26.00
[5,]	5	23	123	30	46.25
[6,]	6	36	216	42	175.00

Expt. No. _____

Experiment -3.

Aim:

Write a program in R to execute the following:

A) Use Readfile and Dataframe to store contents in R data structure.

B) Compute the total subject Marks for each student.

C) Find the median and average of subject 1, subject 2, subject 3, subject 4, subject 5 for each students.

D) Find the Average of the total marks.

E) Plot a line plot of average marks for each student.

F) Add Median of subject score for each student to the line plot

G) Find the list of students who have scored more than 90 in total marks.

F) Find list of students with DOB greater than 1-07-1940 & status Action.

I) Find Interquartile Deviation for Total Marks column.

J) Compare the performance

Teacher's Signature _____

Expt. No. _____ Page No. _____

K) Put the dataframe with additional column back to end and save.

Code:

A) df = read.csv("\\Users\\Downloads\\ testfile.csv")

B) df $sum = rowSums(df[, c[3:]])

C) df $mean = rowMeans(df, c[3:7])
df $median = apply(df[, 3:]], 1, median)

D) mean(sum)

E) plot(as. factor[, p], cr, type = "l", col = 'Red")

F) df $ total sum Arg = (1. rowSums(df, [, c[3:7])

G) print(total sum Arg)

H) df $$subject. median = apply(df[, 3:7], 1, median, na.rm = TRUE)

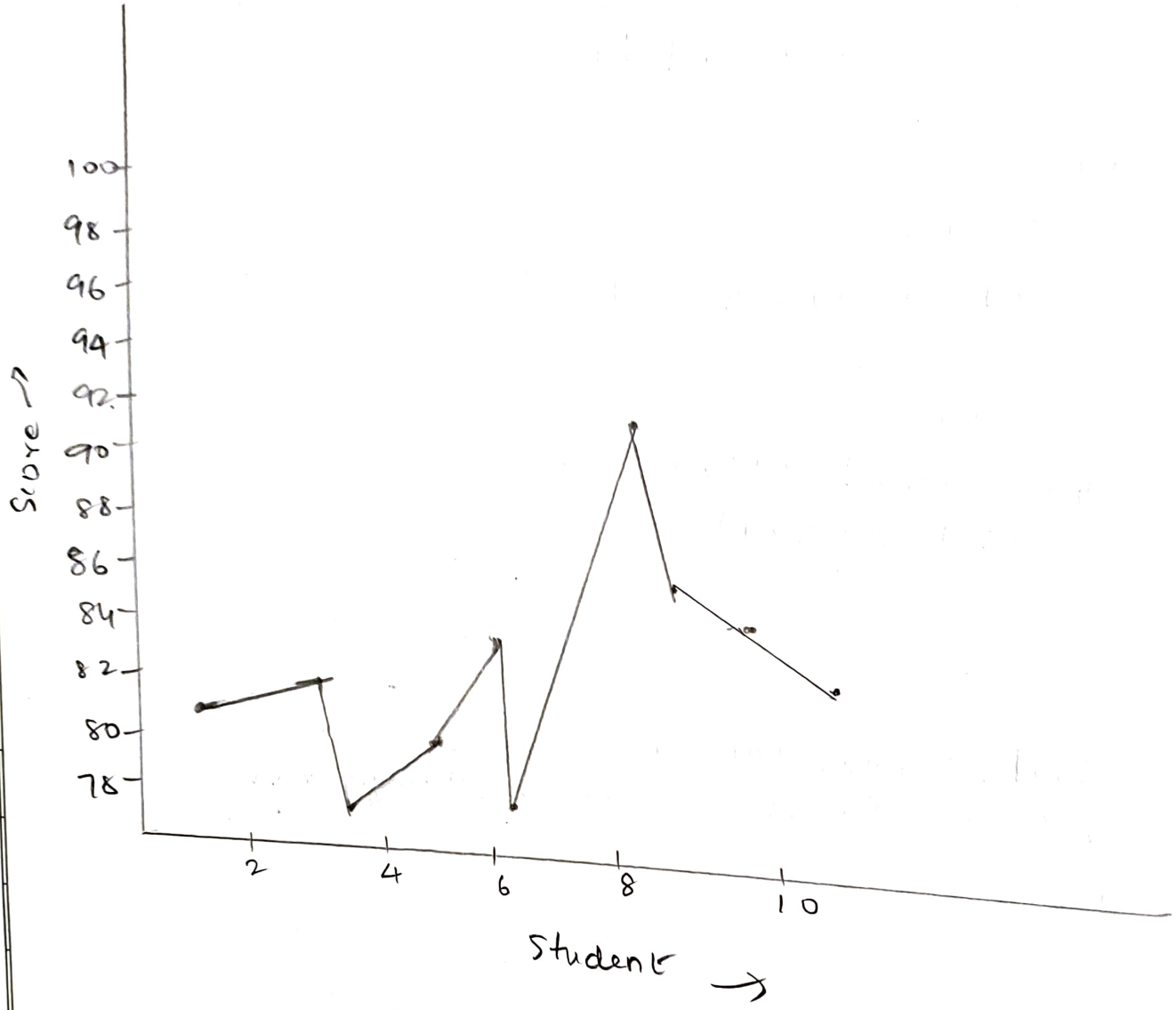
I) print(subject median)

J) The barplot describes a class performance i increase in the metrics of marks while that of 1st, 5th and 7th student remaining consisting high.

K) write csv(df, "Users\\ Downloads\\ testfile.csv")

Teacher's Signature _____

E).



Expt. No. _____

Experiment-4

Aim:

For the given spreadsheet, perform following operations.

1) Read both files.
2) Add a column to compute net revenue price for sugar
3) Add a column to compute net revenue price for maize
4) Plot a histogram of production in tonns both for sugar and maize.
5) Find average, median of revenue both for sugar and maize.
6) Perform a linear correlation of the revenues of sugar and Maize. Make inferences on the obtained values. Is linear correlation valid?
7) Perform a linear regression in one variable of revenue vs maize production in tonnes.
8) Perform a linear regression in one variable of revenue vs sugar production in tonnes
9) Obtain intercept & slope for both the lines and plot J against n for cases 7 & 8.

Code:

```
df ← read.csv("____")
print(df)
df1 ← read.csv("____")
print(df1)
```

Teacher's Signature _____

Page No.

Expt. No.

```
df $revenue = (df[,1] * df[,2] + 0.1 [df[,1] * df[,2])
df1 $revenue = (df1[,1] * df[,2] + 0.5r * (df1[,1] * df[,2]))

arg = df_arg = mean (df[,2])
median = df_median (df[,2])
df1_arg = mean (df[,2])
df1_arg
df1_median = median (df1[,2])
df1_median

df1_rerarg = mean (df1[,5])
df1_rerarg

df1_rermed = median (df1 c[,5])
df1_rermed.

df_rerarg = mean (df[,5])
df_rerarg
df_rermed = median (df[,5])
df_rermed

result = con ([df[,5], df1[,5], method = "Pearson"]
result

ln maize ← ln (df1[,5] ~ df[,2])
ln maize
```

Teacher's Signature

28

Experiment-5

Aim:

Write a program in R to execute the following:

1) Read Json file (labtask5) using R data frame.
2) Find average salary, standard deviation, median for all employees.
3) Project 2 year salary for each employee where salary increase 10% on an average every year.
4) Express earnings as a percentage of salary for each employee.
5) Plot a linear line of Earning vs Salary.
6) Perform a linear regression of Employee Salary.
7) Predict earnings for each of employee salary.
8) Record differences in predicted earnings vs actual earnings.

Code:

```
install.packages("rjson")
library("rjson")
input <- fromJSON(file="/Users/Manasa/Downloads/labtask5.json")
df = as.data.frame(input)

df$ID <- as.numeric(sapply(strsplit(as.character(df$ID),","),
          function(x) x[1]))
df$Salary <- as.numeric(sapply(strsplit(as.character(df$Salary),","),
          function(x) x[1]))
df$Earnings <- as.numeric(sapply(strsplit(as.character(df$Earnings),","),
          function(x) x[1]))
df
```

Teacher's Signature

Expt. No. _____ Page No. _____

```
mean(df$Salary, na.rm = TRUE)
sd(df$Salary, na.rm = TRUE)
median(df$Salary, na.rm = TRUE)

df$OneYearSalary = (df[,2] + (df[,2] * 0.1))
df
df$TwoYearSalary = (df[,6] + (df[,6] * 0.1))
df

for(x in 1:25) {
  print(paste("Earning as Percentage of Salary For Employees", x,
  "=", (df$Earnings[x] / df$Salary[x]) * 100, "%"))
}

x <- sort(x)
y <- sort(y)
plot(x, y, type = "l", xlab = "Salary", ylab = "Earnings")

plot(x, y, col = "blue", xlab = "Salary", ylab = "Earnings", main = "Earnings
     vs Salary Regression")
abline(lm(y ~ x), col = "Red")

model <- lm(y ~ x, data = df)
model
predict(model, newdata = df)

df$Predicted <- predicted(model, newdata = df)
df$Difference <- (df[,8] - df[,3])
```

Teacher's Signature _____ 20/10/2

Expt. No. _____

Experiment-6

Aim:
Write a program in R to execute the following:

1. library(datasets)
   str(iris)

2. Get first 10 rows of each subset.
   → subset(iris, Species == "setosa")[1:10,]
   subset(iris, Species == "versicolor")[1:10,]
   subset(iris, Species == "virginica")[1:10,]

3. summary(iris)

4. Perform correlation among variables.
   → corr <- cor(iris[, 1:4])
   round(corr, 3)

5. Classification with Decision Trees
   Create a model that predicts the species from the petal & sepal width and length. One model that is easy to create and understand is a decision tree, which can be created with C5.0 package.

   → install.packages('C50')
   library(C50)

6. model1 <- c5.0(input, output, control = C5.0control(noGlobalPruni
TRUE, minCases=1))

plot(model1, main="C5.0 Decision Tree - Unpruned, min=1")

7. model2 <- c5.0(input, output, control = C5.0control(noGlobalPruning=
FALSE))

plot(model2, main="C5.0 Decision Tree - Pruned")

Question 1: What does step 3 provide?

Step 3 provides a summary of all ~~three~~ species and the factor present in dataset.

2. What is statistical correlation? What can you infer from the results in step 4?

– Statistical correlation express the extent in which two variables are linearly related.

	SL	S.W	PL	PW
Sepal length	1.000	-0.110	0.872	0.818
Sepal Width	-0.118	1.000	-0.428	-0.36
Petal Length	0.872	-0.428	1.000	
Petal Width	0.818	0.366	0.263	1.000

Teacher's Signature _____

Expt. No. _____ Page No. _____

3. What is decision tree?

- Decision tree is used to make effective decision for classification as well as prediction.
Here, it is used to classify a given flower with details of petals and sepal in one of the categories.

4. What can you understand from outcome of step 5?

- Step 5 shows an unpruned tree, therefore tree has 5 nodes.

Teacher's Signature _____

Expt. No. 07

Experiment - 7

Aim:

Write a program in R to execute the following:

1. Load the iris dataset
2. install packages
3. Create GG pairs for the whole iris data set.
4. Create GG pairs for only first 4 columns.
5. Create subset of 2 species by taking first 100 rows. From this 100 randomly take 80 rows as training set & take 20 as test set.
6. Do ggpairs() for both test and training.

Code:

```
library(datasets)
str(iris)
installed.packages()
install.packages("GGally")
install.packages("ggplot2")

library(ggplot2)
library(GGally)
ggpairs(iris)
ggpairs(iris, columns=1:4)
```

Teacher's Signature

```
ir_data = iris[1:100,]
ir_data
sample = sample(c(TRUE, FALSE), nrow(ir_data), replace = TRUE,
                prob = c(0.8, 0.2))
test = ir_data[!sample, ]
train = ir_data[sample, ]
ggpairs(test)
ggpairs(train).
```