**PART-A**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 1**

**1. Write a program to create a 3 X 3 matrices A and B and perform the following operations**

**a. AT.B**

**b. BT.(A.AT)**

**c. (A.AT).BT**

**d. [(B.BT)+(A.AT)-100I3]-1**

**#CODE:**

#CREATEING A MATRIX

A <- matrix(data = c(1,2,3,4,5,6,7,8,9),nrow = 3,ncol = 3,byrow = TRUE)

B <- matrix(data = c(9,8,7,6,5,4,3,2,1),nrow = 3,ncol = 3,byrow = TRUE)

#DEFINE INDENTITY MATRIX I3

I3 <- diag(3)

#1. AT.B

result\_a <- t(A)%\*%B

#2. BT.(A.AT)

result\_b <- t(B)%\*%(A%\*%t(A))

#3. (A.AT).BT

result\_c <- (A%\*%t(A))%\*%t(B)

#4.[(B.BT)+(A.AT)-100\*I3]^-1

result\_d <- solve((B%\*%t(B))+(A%\*%t(A))-100\*I3)^-1

#PRINT THE RESULTS

cat("a. AT.B :\n")

print(result\_a)

cat("b. BT.(A.AT) :\n")

print(result\_b)

cat("c. (A.AT).BT :\n")

print(result\_c)

cat("d. [(B.BT)+(A.AT)-100\*I3]^-1 :\n")

print(result\_d)

**OUTPUT:**

a. AT.B :

[,1] [,2] [,3]

[1,] 54 42 30

[2,] 72 57 42

[3,] 90 72 54

b. BT.(A.AT) :

[,1] [,2] [,3]

[1,] 468 1116 1764

[2,] 372 885 1398

[3,] 276 654 1032

c. (A.AT).BT :

[,1] [,2] [,3]

[1,] 732 444 156

[2,] 1758 1065 372

[3,] 2784 1686 588

d. [(B.BT)+(A.AT)-100\*I3]^-1 :

[,1] [,2] [,3]

[1,] 16.19325 235.0649 -15.81131

[2,] 235.06494 -174.0385 235.06494

[3,] -15.81131 235.0649 16.19325

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 2**

**Write R program to find roots of quadratic equation using user defined function. Test the program user supplied values for all possible cases.**

**#CODE:**

quadratic\_roots <- function(a, b, c) {

# Calculate the discriminant

discriminant <- b^2 - 4 \* a \* c

# Check if the discriminant is positive, negative, or zero

if (discriminant > 0){

# Two real and distinct roots

root1 <- (-b + sqrt(discriminant)) / (2 \* a)

root2 <- (-b - sqrt(discriminant)) / (2 \* a)

cat("Two real and distinct roots:\n")

cat("Root 1:", root1, "\n")

cat("Root 2:", root2, "\n")

} else if (discriminant == 0) {

# One real root (repeated)

root <- -b / (2 \* a)

cat("One real root (repeated):\n")

cat("Root:", root, "\n")

} else {

# Complex roots

real\_part <- -b / (2 \* a)

imaginary\_part <- sqrt(-discriminant) / (2 \* a)

cat("Complex roots:\n")

cat("Root 1:", real\_part, "+", imaginary\_part, "\n")

cat("Root 2:", real\_part, "-", imaginary\_part, "\n")

}

}

# Prompt the user to enter coefficients

cat("Enter the coefficients of the quadratic equation:\n")

a <- as.numeric(readline(prompt = "a: "))

b <- as.numeric(readline(prompt = "b: "))

c <- as.numeric(readline(prompt = "c: "))

# Call the quadratic\_roots function with user-supplied coefficients

quadratic\_roots(a, b, c)

**OUTPUT :**

a: 4

b: 5

c: 6

Complex roots:

Root 1: -0.625 + 1.053269

Root 2: -0.625 - 1.053269

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 3**

**Write R script to generate prime numbers between two numbers using loops.**

**#CODE:**

#Function to check if a number is prime

is\_prime <- function(n){

if(n<=1){

return(FALSE)

}

if(n<=3){

return(TRUE)

}

if(n%%2==0){

return(FALSE)

}

i <- 5

while(i\*i<=n){

if(n%%i==0||n%%(i+2)==0){

return(FALSE)

}

i <- i+6

}

return(TRUE)

}

#Function to generate prime numbers between two numbers

generate\_primes <- function(start,end){

prime <- c()

for(i in start:end){

if(is\_prime(i)){

prime <- c(prime,i)

}

}

return(prime)

}

#input start and end values from the user

start\_num <- 5

end\_num <- 50

#generate prime numbers between start\_num and end\_num

prime\_numbers <- generate\_primes(start\_num,end\_num)

#print the prime number

cat("Prime Numbers Between",start\_num,"and",end\_num,"are",prime\_numbers,"\n")

**OUTPUT :**

**Prime Numbers Between 5 and 50 are 5 7 9 11 13 15 17 19 21 23 27 29 31 33 37 39 41 43 47**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 4**

**Write an R program to create a list containing strings, numbers, vectors and logical values and do the following manipulations over the list**

1. **Access the first element in the list**
2. **Give the names to the elements in the list**
3. **Add element at some positions in the list**
4. **Remove the element**
5. **print the first and third element**
6. **Update the third element**

**#CODE:**

#CREATE A LIST WITH DIFFERENT DIFFERENT TYPES OF ELEMENTS

my\_list <- list("apple",42,c(1,2,3),TRUE)

#A.ACCESS THE FIRST ELEMENT IN THE LIST

first\_element <- my\_list[[1]]

cat("a.First Element In The List :",first\_element,"\n")

#B.GIVE NAME TO THE ELEMENTS OF THE LIST

names(my\_list) <- c("fruits","numbers","vector","logical")

#C.ADD AN ELEMENT AT SOME POSITIONS IN THE LIST

my\_list[["new\_elements"]] <- "banana"

#D.REMOVE AN ELEMENT

if("number"%in%names(my\_list)){

my\_list[["number"]] <- NULL

}

#E. PRINT THE FIRST AND THIRD ELEMENTS

first\_element <- my\_list[["fruits"]]

third\_element <- my\_list[["vector"]]

cat("First Element In The List :",first\_element,"\n")

cat("Third Element In The List :",third\_element,"\n")

#F.UPDATE THE THIRD ELEMENT

my\_list[["vector"]] <- c(4,5,6)

#PRINT THE UPDATED LIST

cat("Updated list :\n")

print(my\_list)

**OUTPUT :**

**First Element In The List : apple**

**Third Element In The List : 1 2 3**

**Updated list :**

**$fruits**

**[1] "apple"**

**$numbers**

**[1] 42**

**$vector**

**[1] 4 5 6**

**$logical**

**[1] TRUE**

**$new\_elements**

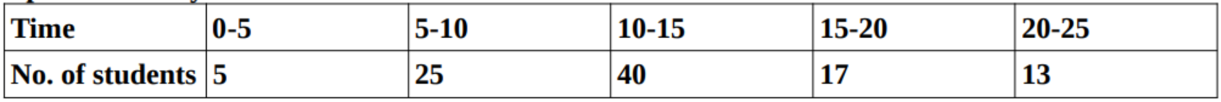
**[1] "banana"**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 5**

**The following table shows the time taken (in minutes) by 100 students to travel to school on a particular day.**

****

1. **Draw the histogram**
2. **Draw frequency polygon**

**#CODE:**

#SAMPLE DATA(REPLACE THIS WITH YOUR ACTUAL DATA)

time\_taken <- c(5,10,15,20,25)

no\_of\_students <- c(5,25,40,17,13)

#CREATE A HISTOGRAM

barplot(no\_of\_students,names.arg=time\_taken,main="Histogram Of Time Taken To Travel To School", xlab="Time Taken(Minutes)",ylab="Number Of Students",col="blue")

#CREATE A FREQUENCY POLYGON

plot(time\_taken, no\_of\_students, type="o", main="Frequency Polygon Of Time Taken To Travel To School", xlab = "Time Taken(Minutes)",ylab="Number Of Students",col="red")

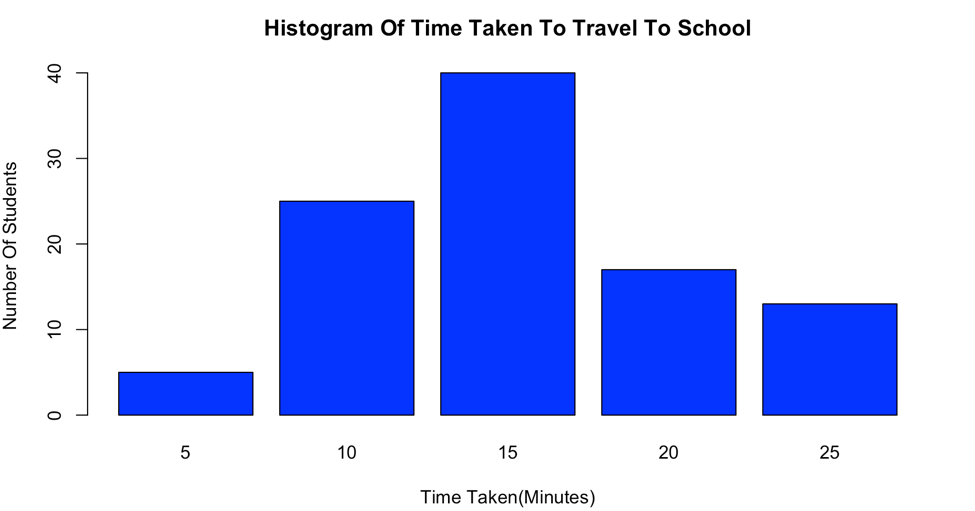
#OPTIONALLY, ADD POINTS FOR CLARITY

points(time\_taken, no\_of\_students, pch=19, col="red")

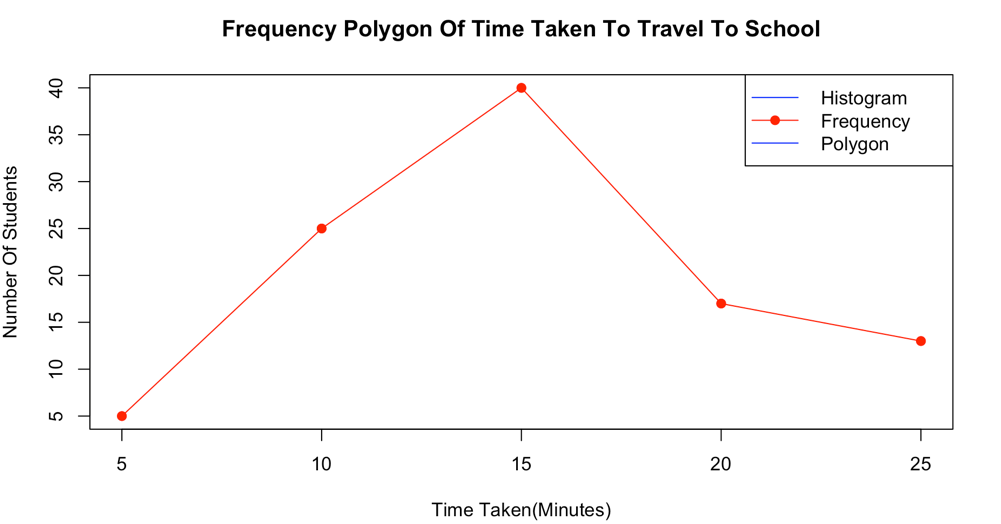
#OPTIONALLY, ADD A LEGEND

legend("topright", legend=c("Histogram","Frequency","Polygon"),col = c("blue","red"),lty = 1,pch = c(NA,19))

**OUTPUT:**

****

1. **The histogram**

****

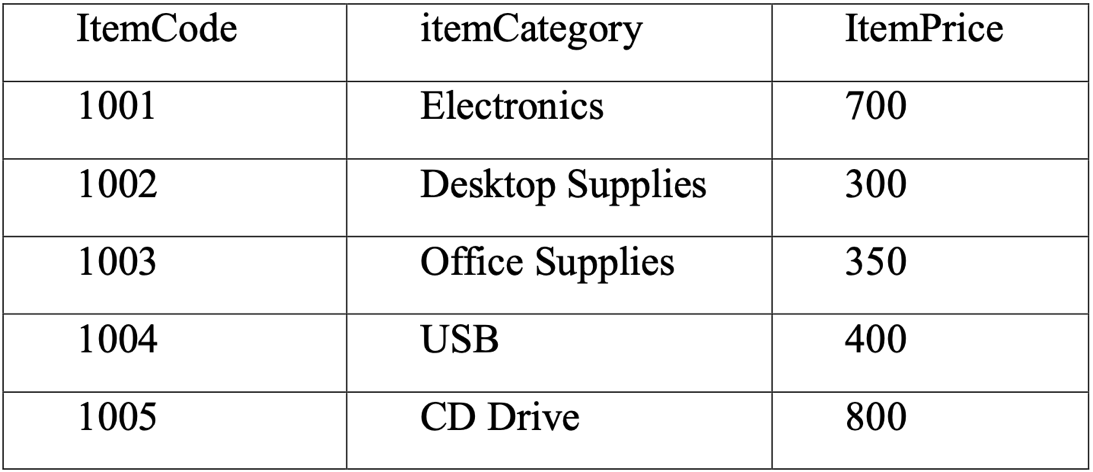
1. **Frequency polygon**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 6**

**Write an R program to create a Data Frame with following details and do the following operations.**

****

1. **Subset the Data frame and display the details of only those items whose price is greater than or equal to 350.**
2. **Subset the Data frame and display only the items where the category is either “Office Supplies” or “Desktop Supplies”**
3. **Subset the Data frame and display the items where the Itemprice between 300 and 700**
4. **Compute the sum of all ItemPrice**
5. **Create another Data Frame called “item-details” with three different fields itemCode, ItemQtyonHand and ItemReorderLvl and merge the two frames.**

**#CODE:**

#ORIGNAL DATA FRAME

data <- data.frame(ItemCode=c(1001,1002,1003,1004,1005),ItemCategory=c("electronics","Desktop Suppliers","Office Suppliers","USB","CD Drive"),ItemPrice = c(7000,300,350,400,800))

#A. SUBSET ITEMS WITH PRICE >=350

subset\_a <- data[data$ItemPrice>=350,]

cat("Item with Price >= 350 :\n")

print(subset\_a)

#B. SUBSET ITEMS WITH CATEGORY "OFFICE SUPPLIERS" OR "DESKTOP SUPPLIES"

subset\_b <- data[data$ItemCategory%in%c("Offfice Supplies","Desktop Supplies"),]

cat("Items With Category 'Office Supplies' or 'Desktop Supplies' :\n")

print(subset\_b)

#C. SUBSET WITH ITEMPRICE BETWEEN 300 AND 700

subset\_c <- data[data$ItemPrice>=300&data$ItemPrice<=700,]

cat("Items With Price Between 300 and 700 :\n")

print(subset\_c)

#d.COMPUTE THE SUM OF ALL ITEM PRICE

total\_price <- sum(data$ItemPrice)

cat("total Item Price :",total\_price,"\n")

#E. CREATE ANOTHE FRAME CALLED "ITEM-DETAILS"

item\_details <- data.frame(ItemCode=c(1001,1002,1003,1004,1005),ItemQtyOnHand=c(10,20,15,30,25),ItemReOrderLvl=c(5,10,8,15,12))

#MERGE THE TWO DATA FRAME ON THE "ITEMCODE" COLUMN

merge\_data <- merge(data,item\_details,by = "ItemCode")

print(merge\_data)

**OUTPUT:**

**Item with Price >= 350 :**

**ItemCode ItemCategory ItemPrice**

**1 1001 electronics 7000**

**3 1003 Office Suppliers 350**

**4 1004 USB 400**

**5 1005 CD Drive 800**

**Items With Category 'Office Supplies' or 'Desktop Supplies' :**

**[1] ItemCode ItemCategory ItemPrice**

**<0 rows> (or 0-length row.names)**

**Items With Price Between 300 and 700 :**

**ItemCode ItemCategory ItemPrice**

**2 1002 Desktop Suppliers 300**

**3 1003 Office Suppliers 350**

**4 1004 USB 400**

**total Item Price : 8850**

**data.frame(ItemCode=c(1001,1002,1003,1004,1005),ItemQtyOnHand=c(10,20,15,30,25),ItemReOrderLvl=c(5,10,8,15,12))**

**ItemCode ItemCategory ItemPrice ItemQtyOnHand ItemReOrderLvl**

**1 1001 electronics 7000 10 5**

**2 1002 Desktop Suppliers 300 20 10**

**3 1003 Office Suppliers 350 15 8**

**4 1004 USB 400 30 15**

**5 1005 CD Drive 800 25 12**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 7**

**Create a factor marital\_status with levels Married, single, divorced. Perform the**

**following operations on this factor**

1. **Check the variable is a factor**
2. **Access the 2nd and 4th element in the factor**
3. **Remove third element from the factor**
4. **Modify the second element of the factor**
5. **Add new level widowed to the factor and add the same level to the factor marital\_status**

**#CODE:**

#CREATE THE FACTOR VARIABLE MARITAL\_STATUS

marital\_status <- factor(c("Married","Single","Adult","Divorced","child"))

#A.CHECK IF THE VARIABLE IS A FACTOR

is.factor(marital\_status)

#B. ACCESS THE 2ND AND 4TH ELEMANT IN THE FACTOR

second\_element <- marital\_status[2]

second\_element

fourth\_element <- marital\_status[4]

fourth\_element

#C.REMOVE THE THIRD ELEMENT FROM THE FACTOR

marital\_status <- marital\_status[-3]

print(marital\_status)

#D.MODIFY THE SECOND ELEMENT OF THE FACTOR

marital\_status[2] <- "child"

marital\_status

#E.ADD A NEW LEVEL "WIDOWED" TO THE FACTOR

marital\_status <- factor(marital\_status,levels = c(levels(marital\_status),"Widowed"))

#ADD THE SAME LEVEL "WIDOWED" TO THE FACTOR

marital\_status[3] <- "Widowed"

#VIEW THE MODIFIED FACTOR

marital\_status

**OUTPUT:**

**A. CHECK IF THE VARIABLE IS A FACTOR**

**[1] TRUE**

**B. ACCESS THE 2ND AND 4TH ELEMANT IN THE FACTOR**

**[1] Single**

**Levels: Adult child Divorced Married Single**

**[1] Divorced**

**Levels: Adult child Divorced Married Single**

**C. REMOVE THE THIRD ELEMENT FROM THE FACTOR**

**[1] Married Single Divorced child**

**Levels: Adult child Divorced Married Single**

**D. MODIFY THE SECOND ELEMENT OF THE FACTOR**

**[1] Married child Divorced child**

**Levels: Adult child Divorced Married Single**

**E. ADD A NEW LEVEL "WIDOWED" TO THE FACTOR AND ADD THE SAME LEVEL "WIDOWED" TO THE FACTOR AND THNVIEW THE MODIFIED FACTOR**

**[1] Married child Widowed child**

**Levels: Adult child Divorced Married Single Widowed**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 8**

**Write a R language Script for following operation on Iris Data Set**

1. **Load the Iris Dataset**
2. **View first six rows of iris dataset**
3. **Summarize iris dataset**
4. **Display number of rows and columns**
5. **Display column names of dataset.**
6. **Create histogram of values for sepal length**
7. **Create scatterplot of sepal width vs. sepal length**
8. **Create boxplot of sepal width vs. sepal length**
9. **Find Pearson correlation between Sepal.Length and Petal.Length**
10. **Create correlation matrix for dataset.**

**#CODE:**

#LOAD THE IRIS DADASET

data(iris)

#VIEW THE FIRST SIX ROWS OF THE IRIS DATASET

head(iris)

#SUMMARISE THE IRIS DATASET

summary(iris)

#DISPLAY THE NUMBER OF ROWS AND COLUMS

dim(iris)

#DISPLAY COLUMN NAMES

colnames(iris)

#CREATE A HISTOGRAM OF SEPAL LENGTH

hist(iris$Sepal.Length,main="Histogram of Sepal Length",xlab = "Sepal Length")

#CREATE A SCATTERPLOT OF SEPAL WIDTH VS. SEPAL LENGTH

plot(iris$Sepal.Length,iris$Sepal.Width,main="Scatterplot of Sepal Width vs. Sepal Length",xlab = "Sepal Length",ylab = "Sepal Width")

#CREATE A BOXPLOT OF SEPAL WITH VS.SEPAL LENGTH

boxplot(iris$Sepal.Width ~ iris$Sepal.Length,main="Boxplot of Sepal Width vs.Sepal Length",xlab = "Sepal Length",ylab = "Sepal Width")

#FIND PEARSON CORRELATION BETWEEN SEPAL.LENGTH AND PETAL.LENGTH

cor(iris$Sepal.Length,iris$Petal.Length,method = "pearson")

#CREATE A CORELATION MATRIX FOR THE ENTIRE DATASET

cor(iris[,1:4])

**OUTPUT:**

**THE FIRST SIX ROWS OF THE IRIS DATASET**

**Sepal.Length Sepal.Width Petal.Length Petal.Width Species**

**1 5.1 3.5 1.4 0.2 setosa**

**2 4.9 3.0 1.4 0.2 setosa**

**3 4.7 3.2 1.3 0.2 setosa**

**4 4.6 3.1 1.5 0.2 setosa**

**5 5.0 3.6 1.4 0.2 setosa**

**6 5.4 3.9 1.7 0.4 setosa**

**SUMMARISED THE IRIS DATASET**

**Sepal.Length Sepal.Width Petal.Length Petal.Width Species**

**Min. :4.300 Min. :2.000 Min. :1.000 Min. :0.100. setosa :50**

**1st Qu.:5.100 1st Qu.:2.800 1st Qu.:1.600 1st Qu.:0.300 versicolor:50 Median :5.800 Median :3.000 Median :4.350 Median :1.300 virginica :50**

**Mean :5.843 Mean :3.057 Mean :3.758 Mean :1.199**

**3rd Qu.:6.400 3rd Qu.:3.300 3rd Qu.:5.100 3rd Qu.:1.800**

**Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500**

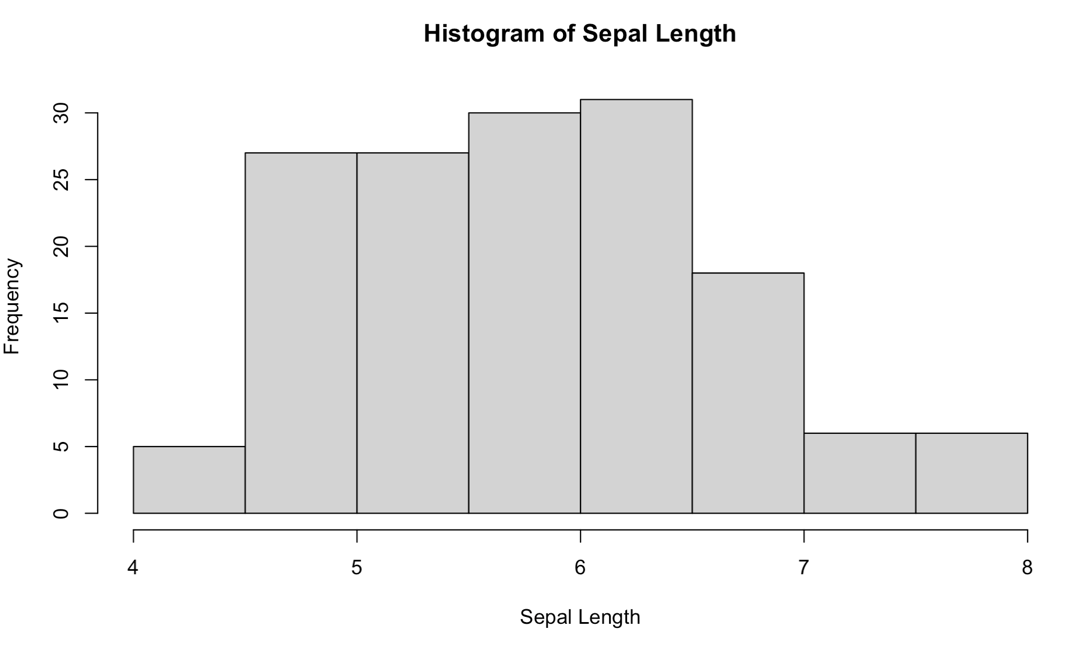
**THE NUMBER OF ROWS AND COLUMS**

**[1] 150 5**

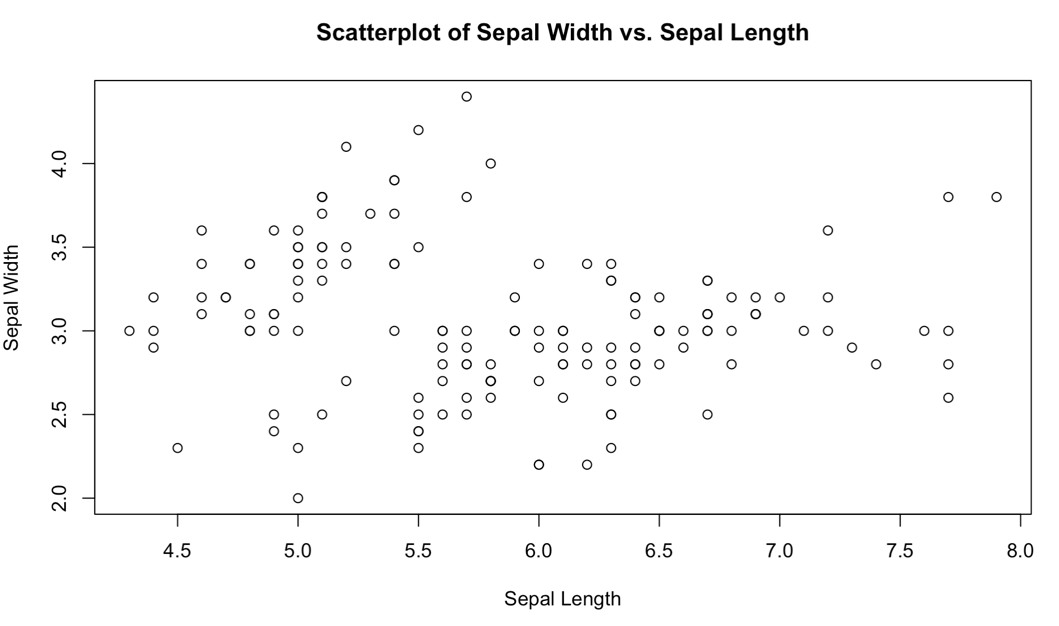
**COLUMN NAMES**

**[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"**

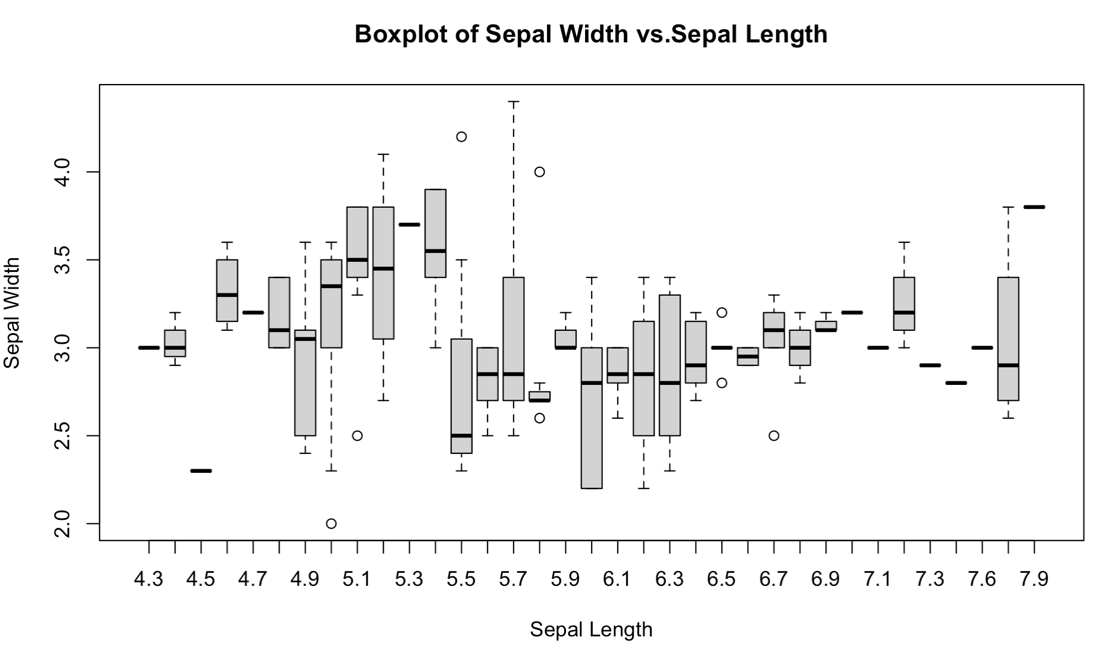
**CREATE A HISTOGRAM OF SEPAL LENGTH**



SCATTERPLOT OF SEPAL WIDTH VS. SEPAL LENGTH



BOXPLOT OF SEPAL WITH VS.SEPAL LENGTH



**PEARSON CORRELATION BETWEEN SEPAL.LENGTH AND PETAL.LENGTH**

**[1] 0.8717538**

**CREATE A CORELATION MATRIX FOR THE ENTIRE DATASET**

**Sepal.Length Sepal.Width Petal.Length Petal.Width**

**Sepal.Length 1.0000000 -0.1175698 0.8717538 0.8179411**

**Sepal.Width -0.1175698 1.0000000 -0.4284401 -0.3661259**

**Petal.Length 0.8717538 -0.4284401 1.0000000 0.9628654**

**Petal.Width 0.8179411 -0.3661259 0.9628654 1.0000000**

**PART-B**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 1**

**Write a R program to create a Vector containing following 8 values and perform the following operations.**

**4 3 0 5 2 9 4 5**

1. **Find mean, median, mode.**
2. **Find the range.**
3. **Find the 35th and 78th percentile.**
4. **Find the variance and standard deviation**
5. **Find the interquartile range.**
6. **Find the z-score for each value.**

**#CODE:**

#CREATE A VECTOR WITH THE GIVEN VALUES

values <- c(4,3,0,5,2,9,4,5)

#A.FIND MEAN,MEADIAN AND MODE

mean\_value <- mean(values)

median\_values <- median(values)

frequency\_table <- table(values)

max\_freq <- max(frequency\_table)

mode\_values <- as.numeric(names(frequency\_table[frequency\_table==max\_freq]))

#B.FIND THE RANGE

range\_values <- max(values)-min(values)

#C.FIND THE 34TH AND 78TH PERCENTILE

percentile\_35 <- quantile(values,0.35)

percentile\_78 <- quantile(values,0.78)

#D.FIND THE VARIENCE AND STANDARD DEVIATION

variance\_value <- var(values)

std\_deviation\_value <- sd(values)

#E.FIND THE INTERQUARTILE RANGE

quantiles <- quantile(values,probs=c(0.25,0.75))

interquartile\_range <- diff(quantiles)

#F.FIND THE Z-SCORE FOR EACH VALUE

z\_score <- (values-mean\_value)/std\_deviation\_value

#PRINT THE RESULTS

cat("A.\t Mean :",mean\_value,"\n\t Median :",median\_values,"\n\t Mode :",mode\_values,"\n\n")

cat("B.\t Range :",range\_values,"\n\n")

cat("C.\t 35th Persentle :",percentile\_35,"\n\t 78th Percentile :",percentile\_78,"\n\n)")

cat("D.\t Varience :",variance\_value,"\n\t Standard Deviation :",std\_deviation\_value,"\n\n")

cat("E.\t Interquartile Range :",interquartile\_range,"\n\n")

cat("F.\t Z-Scores :\n\t",z\_score)

**OUTPUT:**

**A. Mean : 4**

**Median : 4**

**Mode : 4 5**

**B. Range : 9**

**C. 35th Persentle : 3.45**

**78th Percentile : 5**

**D. Varience : 6.857143**

**Standard Deviation : 2.618615**

**E. Interquartile Range : 2.25**

**F. Z-Scores :**

**0 -0.3818813 -1.527525 0.3818813 -0.7637626 1.909407 0 0.3818813**

**MANUALLY SOLVED :**

1. **Mean :**

**Median :Array In Ascending Order**

**Mode :**

1. **Range :**

1. **35th Percentile :**

**78th Percentile :**

1. **Variance :**

|  |  |  |  |
| --- | --- | --- | --- |
| ***x*** | ***x- μ*** | ***| x- μ |*** | ***( x- μ )2*** |
| **0** |  |  |  |
| **2** |  |  |  |
| **3** |  |  |  |
| **4** |  |  |  |
| **4** |  |  |  |
| **5** |  |  |  |
| **5** |  |  |  |
| **9** |  |  |  |
|  |  |  |  |

**Standard Deviation :**

1. **Interquartile Range:**
2. **Z-Score :**

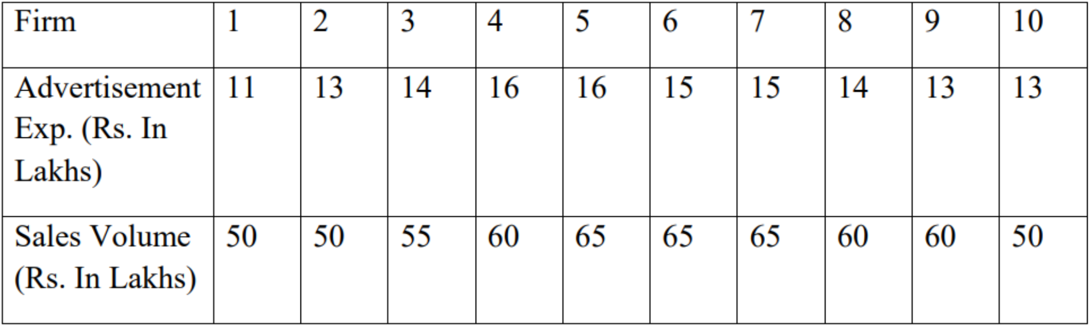
|  |  |  |  |
| --- | --- | --- | --- |
| ***x*** | ***x- μ*** | ***(x- μ)2*** | ***z = ( x - μ ) / σ*** |
| **0** |  |  |  |
| **2** |  |  |  |
| **3** |  |  |  |
| **4** |  |  |  |
| **4** |  |  |  |
| **5** |  |  |  |
| **5** |  |  |  |
| **9** |  |  |  |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 2**

**Write R script to find the correlation coefficient and type of correlation between advertisement expenses and sales volume using Karl Pearson’s coefficient of correlation method (Direct Method).**

****

**#CODE:**

#DEFINE THE DATA

advertisement\_expenses <- c(11,13,14,16,16,15,15,14,13,13)

sales\_volume <- c(50,50,55,60,65,65,65,60,60,50)

#CALCULATE THE MEAN OF ADVERTISEMENT EXPENSES AND SALES VOLUME

mean\_advertisement <- mean(advertisement\_expenses)

mean\_sales <- mean(sales\_volume)

#CALCULATE THE SUM OF PRODUCTS AND SQUARE

sum\_products <- sum((advertisement\_expenses-mean\_advertisement)\*(sales\_volume-mean\_sales))

sum\_square\_advertisement <- sum((advertisement\_expenses-mean\_advertisement)^2)

sum\_square\_sales <- sum((sales\_volume-mean\_sales)^2)

#CALCULATE THE CORRELATION COEFFICIENT

correlation\_coefficent <- sum\_products/(sqrt(sum\_square\_advertisement)\*sqrt(sum\_square\_sales))

#DETERMINE THE TYPE OF CORRELATION

if(correlation\_coefficent>0){

correlation\_type <- "Positive"

}else if(correlation\_coefficent<0){

correlation\_type <- "Nagetive Correlation"

}else{

correlation\_type <- "No correlation(or very weak Correlation)"

}

#PRINT THE RESULS

cat("Correlation Coefficient :",correlation\_coefficent,"\n")

cat("Correlation Type :",correlation\_type,"\n")

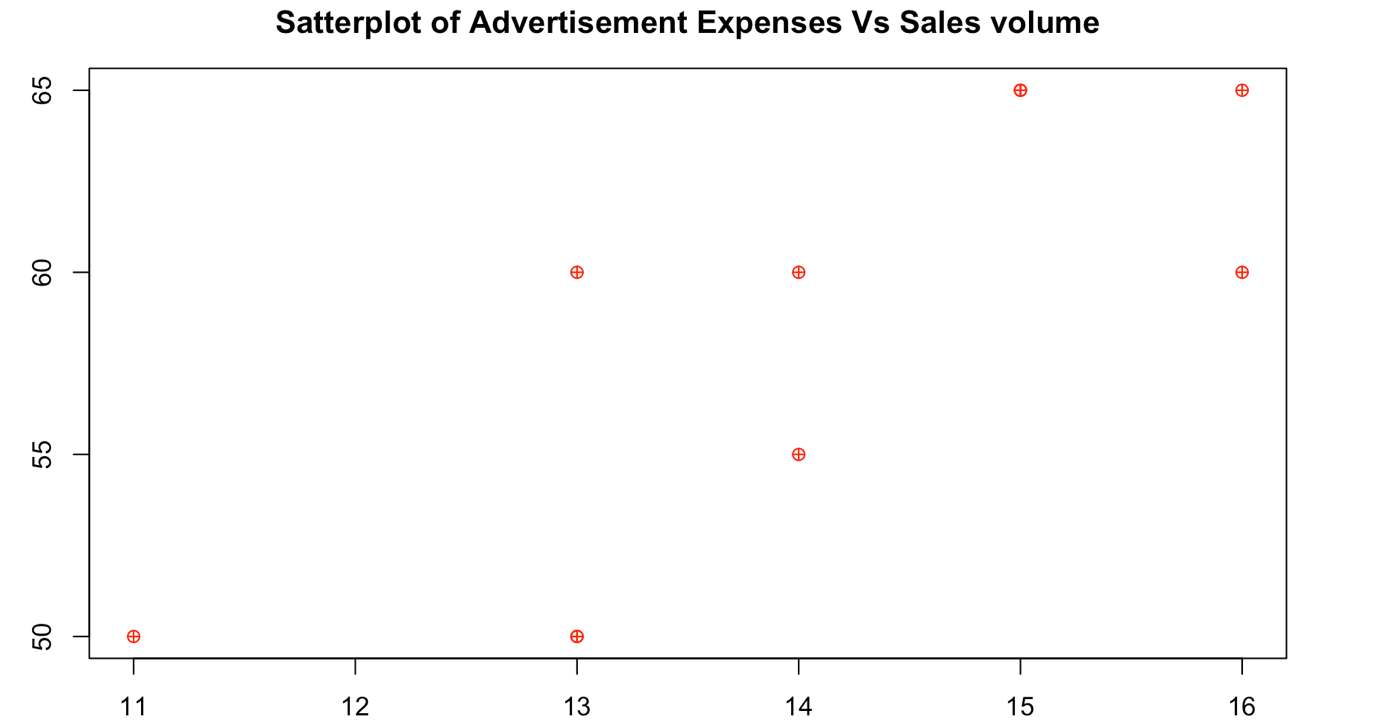
par(mar =c(3,3,3,3))

plot(advertisement\_expenses,sales\_volume,xlab="advertisement expenses",ylab="sales volume",main="Satterplot of Advertisement Expenses Vs Sales volume",col="red",pch=10)

**OUTPUT:**

**Correlation Coefficient : 0.7865665**

**Correlation Type : Positive**

****

**MANUALLY SOLVED :**

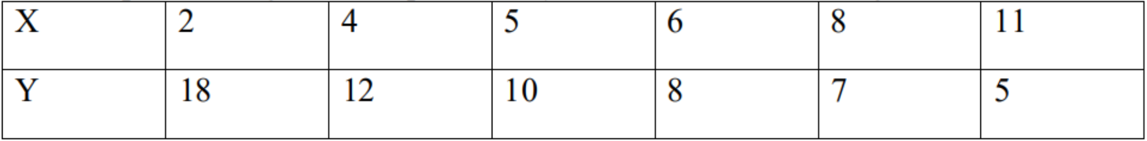
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***x = x - x̄*** | ***x2*** | ***y = y - ȳ*** | ***y2*** | ***x y*** |
| **11** | **50** |  |  |  |  |  |
| **13** | **50** |  |  |  |  |  |
| **14** | **55** |  |  |  |  |  |
| **16** | **60** |  |  |  |  |  |
| **16** | **65** |  |  |  |  |  |
| **15** | **65** |  |  |  |  |  |
| **15** | **65** |  |  |  |  |  |
| **14** | **60** |  |  |  |  |  |
| **13** | **60** |  |  |  |  |  |
| **13** | **50** |  |  |  |  |  |
|  |  |  |  |  |  |  |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 3**

**Write R script to compute the regression equation of y on x from the following data. Predict the value of y when x=7**

****

**#CODE:**

#SAMPLE DATA FOR X AND Y

x <- c(2,4,5,6,7,8)

y <- c(18,12,10,8,7,5)

#FIT A LINEAR REGRESSION MODEL

model <- lm(y~x)

#PRINT THE REGRESSION EQUATION

cat("Regression Equation :Y =",round(coef(model)[1],2),"+",round(coef(model)[2],2),"x\n")

#PREDICT THE VALUES OF Y WHEN X = 7

x\_new <- 7

y\_predicted <- predict(model,data.frame(x=x\_new))

cat("Predicted y when x=",x\_new,":",round(y\_predicted,2),"\n")

**OUTPUT:**

**Regression Equation :Y = 21.2 + -2.1 x**

**Predicted y when x= 7 : 6.5**

**MANUALLY SOLVED :**

|  |  |  |  |
| --- | --- | --- | --- |
| ***x*** | ***y*** | ***x2*** | ***x y*** |
| **2** | **18** |  |  |
| **4** | **12** |  |  |
| **5** | **10** |  |  |
| **6** | **8** |  |  |
| **8** | **7** |  |  |
| **11** | **5** |  |  |
|  |  |  |  |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 4**

**The times taken by a large group of students to complete a piece of homework, T minutes, are Normally distributed with a mean of 57 minutes and standard deviation of 6.5. Find the probability that the time taken by a random student from the group to complete this homework will be less than 60 minutes.**

**Write R script to Find the probability that the time taken by a random student from the group to complete this homework**

1. **Will be less than 60 minutes**
2. **Between 50 and 80 minutes**

**#CODE:**

#MEAN AND STANDARD DEVIATION OF THE DISTRIBUTION

mean\_time <- 57

std\_dev <- 6.5

#CALCULATE THE PROBABILITY THAT TAKEN IS THAN 60 MINUTES

prob\_less\_than\_60 <- pnorm(60,mean\_time,std\_dev)

#CALCULATE THE PROBABILITY THAT THE TIME TAKEN IS BETWEEN 50 AND 80 MINUTES

prob\_between\_50\_and\_80 <- pnorm(80,mean\_time,std\_dev)-pnorm(50,mean\_time,std\_dev)

#PRINT THE RESULTS

cat("a) Probability that time is less than 60 minutes :",round(prob\_less\_than\_60,4),"\n")

cat("b) Probability that time is between 50 and 80 minutes :",round(prob\_between\_50\_and\_80,4),"\n")

**OUTPUT:**

**a) Probability that time is less than 60 minutes : 0.6778**

**b) Probability that time is between 50 and 80 minutes : 0.859**

**MANUALLY SOLVED :**

1. **Will be less than 60 minutes :**

1. **Probability that time is between 50 and 80 minutes :**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 5**

**Write R script to perform the following using binomial distribution**

**i. If n=4 and p=0.10, find P(x=3)**

**ii. If n=12 and p=0.45, find P(5<=x<=7)**

**#CODE:**

#SCENERIO 1: IF N=4 AND P=0.10, FIND P(X=3)

n1 <- 4

p1 <- 0.10

x1 <- 3

#PROBABILITY OF P(X=3)

prob\_x3 <- dbinom(x1,size=n1,prob=p1)

#PRINT THE RESULT

cat("i. Probability that x=3.",round(prob\_x3,4),"\n")

#SENARIO 2: IF N=12 AND P=0.45, FIND P(5<=X<=7)

n2 <- 12

p2 <- 0.45

x2\_min <- 5

x2\_max <- 7

#PROBABILITY OF P(5<=X<=7)

prob\_5\_to\_7 <- pbinom(x2\_max,size = n2,prob = p2)-pbinom(x2\_min-1,size = n2,prob = p2)

#PRINT THE RESULT

cat("ii. Probability that 5<=x<=7:",round(prob\_5\_to\_7,4),"\n")

**OUTPUT:**

1. **Probability that x=3. 0.0036**
2. **Probability that 5<=x<=7: 0.5838**

**MANUALLY SOLVED :**

**i. If n=4 and p=0.10, find P(x=3)**

**ii. If n=12 and p=0.45, find P(5<=x<=7)**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 6**

**Perform the following using uniform distribution between 200 and 240**

**i. P(x>230)**

**ii. P(205≤x≤220)**

**#CODE:**

#DEFINE THE LOWER AND UPPER BOUNDS OF THE UNIFORM DISTRIBUTION

a <- 200

b <- 240

#I.P(X>230)

probability\_x\_gt\_230 <- 1 - punif (230,min=a,max=b)

#II. P(205<=X<=220)

probability\_x\_between\_205\_and\_220 <- punif(220,min = a,max = b)-punif(205,min = a,max = b)

#PRINT THE RESULTS

cat("i. P(x>230)=",probability\_x\_gt\_230,"\n")

cat("ii. P(205<=x<=220)=",probability\_x\_between\_205\_and\_220,"\n")

**OUTPUT:**

1. **P(x>230)= 0.25**
2. **P(205<=x<=220)= 0.375**

**MANUALLY SOLVED :**

**i. P(x>230) :**

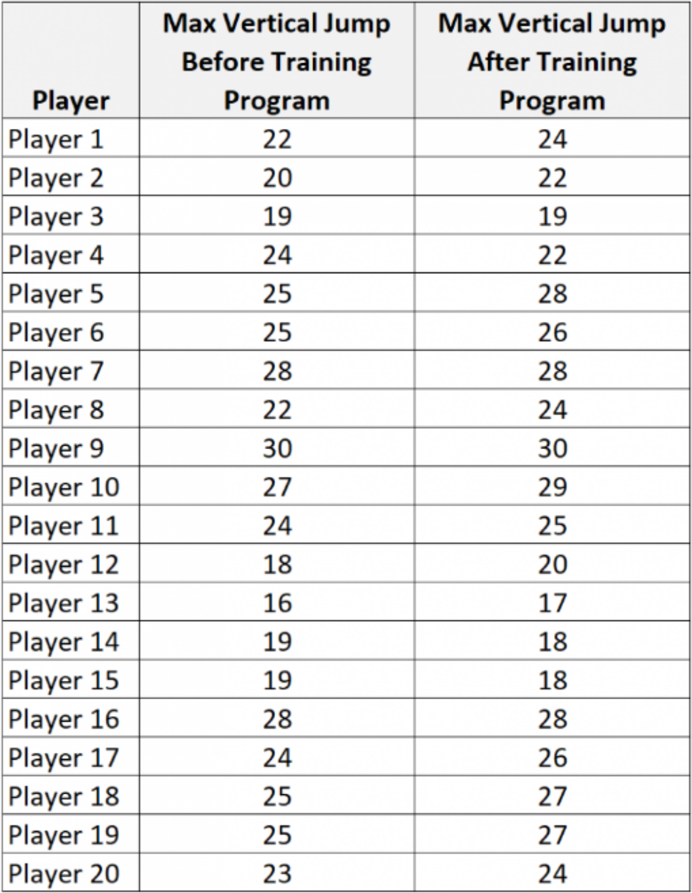
**ii. P(205≤x≤220)**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 7**

**Following are the scores of max vertical jumps before and after the training program. Test whether the training program is helpful to the students (Use Paired t-test).**

****

**#CODE:**

**#CODE:**

#CREATE THE VECTORS WITH SCORES BEFORE AND AFTER THE TRAINING PROGRAM

before <- c(22,20,19,24,25,25,28,22,30,27,24,18,16,19,19,28,24,25,25,23)

after <- c(24,22,19,22,28,26,28,24,30,29,25,20,17,18,18,28,26,27,27,24)

#PERFORM A PAIRED T-TEST

result <- t.test(after,before,paired=TRUE)

#DISPLAY THE RESULTS

cat("paired t-test result :\n")

cat("t-statistic :",result$statistic,"\n")

cat("p-value :",result$p.value,"\n")

**OUTPUT:**

**paired t-test result :**

**t-statistic : 3.226173**

**p-value : 0.004445371**

**MANUALLY SOLVED :**

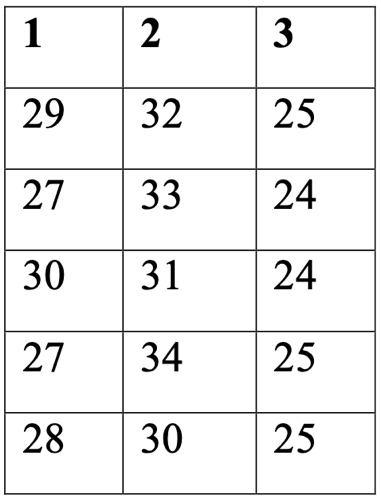
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **PLAYER’S** | **Max Vertical Jump before Training** | **Max Vertical Jump after Training** | **d** | **d-đ** | **( d-đ )2** |
| **Player 1** | **22** | **24** |  |  |  |
| **Player 2** | **20** | **22** |  |  |  |
| **Player 3** | **19** | **19** |  |  |  |
| **Player 4** | **24** | **22** |  |  |  |
| **Player 5** | **25** | **28** |  |  |  |
| **Player 6** | **25** | **26** |  |  |  |
| **Player 7** | **28** | **28** |  |  |  |
| **Player 8** | **22** | **24** |  |  |  |
| **Player 9** | **30** | **30** |  |  |  |
| **Player 10** | **27** | **29** |  |  |  |
| **Player 11** | **24** | **25** |  |  |  |
| **Player 12** | **18** | **20** |  |  |  |
| **Player 13** | **16** | **17** |  |  |  |
| **Player 14** | **19** | **18** |  |  |  |
| **Player 15** | **19** | **18** |  |  |  |
| **Player 16** | **28** | **28** |  |  |  |
| **Player 17** | **24** | **26** |  |  |  |
| **Player 18** | **25** | **27** |  |  |  |
| **Player 19** | **25** | **27** |  |  |  |
| **Player 20** | **25** | **24** |  |  |  |

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*NAME : ABHISHEK KUMAR

REG NO : U05SP21S0009 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**PROGRAM: 8**

**A company has three manufacturing plants, and company officials want to determine whether there is difference in the average age of workers at the three locations. The following data are the ages of five randomly selected workers at each plant. Perform a one-way ANOVA to determine whether there is significant difference in the mean ages of the workers at three plants. Use α=0.01. Write R script for the above problem. Plant(Employee Ages)**

****

**#CODE:**

#DATA

plant\_ages <- matrix(c(29,32,25,27,33,24,30,31,24,27,34,25,28,30,25),nrow = 3,ncol = 5,byrow = TRUE)

plant\_names <- c("Plant 1","Plant 2","Plant 3")

#RESHAPE DATA FOR ANOVA

ages\_vector <- as.vector(plant\_ages)

plant\_vector <- rep(plant\_names,each=5)

#CREATE A DATA FRAME

data\_df <- data.frame(plant=factor(plant\_vector),Age=ages\_vector)

#PERFORM ONE-WAY ANOVA

anova\_result <- aov(Age~plant,data=data\_df)

#CHECK NOVA SUMMARY

summary\_anova <- summary(anova\_result)

#EXRACT P-VALUE

p\_value <- summary\_anova[[1]]$`pr(>F)`[1]

#SIGNIFICANCE LEVEL

alpha <- 0.01

#CHECK IF THERE IS A SIGNIFICANT DIFFRENCE

if(!is.null(p\_value)&&p\_value<alpha){

cat("There is a significant difference in the mean ages of workers at the three plants(reject the null hypothessis).\n")

}

#PRINT ANOVA SUMMARY

cat("\nANOVA Summary :\n")

print(summary\_anova)

**OUTPUT:**

ANOVA Summary :

Df Sum Sq Mean Sq F value Pr(>F)

plant 2 18.53 9.267 0.815 0.466

Residuals 12 136.40 11.367

**MANUALLY SOLVED :**

|  |  |  |
| --- | --- | --- |
| **P1** | **P2** | **P3** |
| **29** | **32** | **25** |
| **27** | **33** | **24** |
| **30** | **31** | **24** |
| **27** | **34** | **25** |
| **28** | **30** | **25** |
|  |  |  |

STEP 1 : F-Table Calculation