

## Experiment1.2

### 1. Aim:

To perform the statistical analysis of data.

### 2. Objective:

Learning about statistical analysis

### 3. Script and Output:

Statistical analysis is a branch of mathematics that deals with the collection, analysis, interpretation, presentation, and organization of data. It involves using various statistical techniques to draw meaningful conclusions from a set of data. Ask or Specify Data Requirements.

The steps involved in performing a statistical analysis are as follows:

1. Define the research question:
2. Collect data
3. Clean and prepare the data
4. Choose an appropriate statistical method
5. Analyse the data
6. Interpret the results

Each step has its own process and tools to make overall conclusions based on the data.

### CODE-

```
library("RWeka")
N=read.arff("super_friends.arff")
# Print Data
print(N)

# Cat is used so that the newline characters are treated as string and not vectors
cat("\n\n\n")

# Printing first two rows from the data set
print(head(N,2))
```



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
# To find the dimensions of data set
dim(N)

# To find the names of variables (Column names) in data set
names(N)

# Show all the friends
N["friends"]

# Show Average school hours
N["avg_school_hours"]

# Show max of average school hours
max(N["avg_school_hours"])

# Show min of average school hours
min(N["avg_school_hours"])

# Sum of average school hours
sum(N["avg_school_hours"])

# Mean of average school hours

#mean(N["avg_school_hours"])

# Create a vector.
x <- c(12,7,3,4.2,18,2,54,-21,8,-5)

# Find Mean.
r<- mean(x)
print(r)

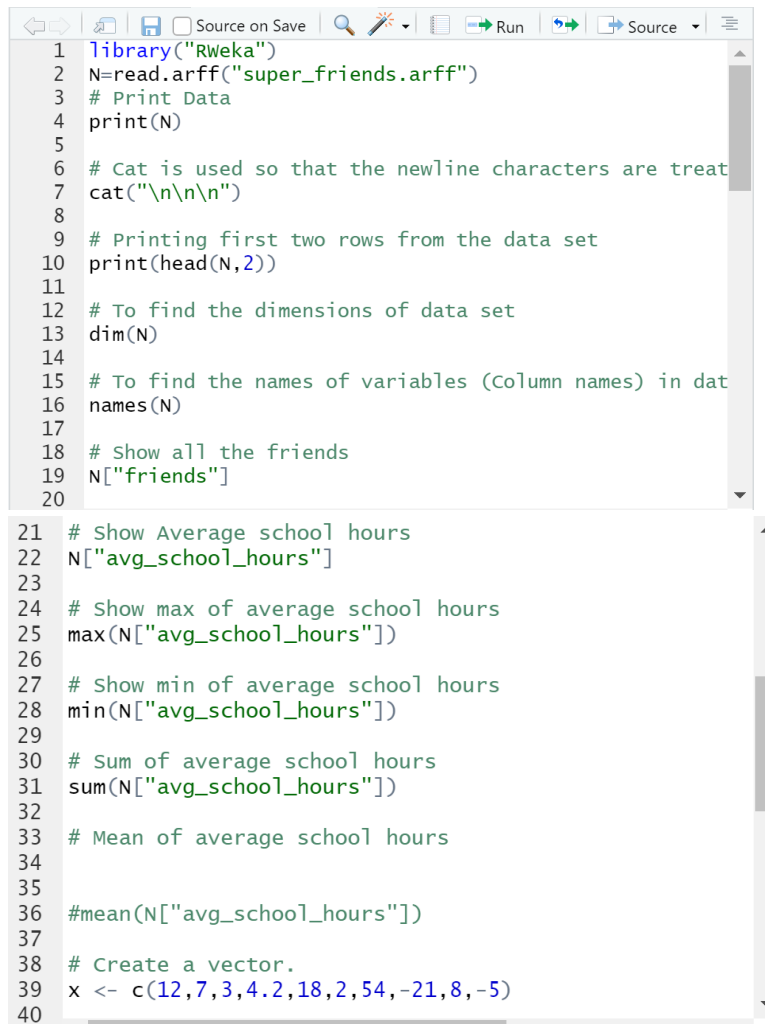
# Median of average school hours
#median(sort(N["avg_school_hours"]))

median(sort(x))

# Standard Deviation of average school hours
#sd(avg_school_hours)

# To generate a summary of data
summary(N)
```

```
level<-c('Topper','Medium','Average','Below Average')
print(class(level))
print(is.factor(level))
levelfact<-factor(level)
print(is.factor(levelfact))
print(levels(levelfact))
print(class(levelfact))
```



```
1 library("Rweka")
2 N=read.arff("super_friends.arff")
3 # Print Data
4 print(N)
5
6 # Cat is used so that the newline characters are treated
7 cat("\n\n\n")
8
9 # Printing first two rows from the data set
10 print(head(N,2))
11
12 # To find the dimensions of data set
13 dim(N)
14
15 # To find the names of variables (Column names) in data set
16 names(N)
17
18 # Show all the friends
19 N["friends"]
20
21 # Show Average school hours
22 N["avg_school_hours"]
23
24 # Show max of average school hours
25 max(N["avg_school_hours"])
26
27 # Show min of average school hours
28 min(N["avg_school_hours"])
29
30 # Sum of average school hours
31 sum(N["avg_school_hours"])
32
33 # Mean of average school hours
34
35
36 #mean(N["avg_school_hours"])
37
38 # Create a vector.
39 x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
40
```

```

50 # Standard Deviation of average school hours
51 #sd(avg_school_hours)
52
53 |
54
55 # To generate a summary of data
56 summary(N)
57
58
59 level<-c('Topper','Medium','Average','Below Average')
60 print(class(level))
61 print(is.factor(level))
62 levelfact<-factor(level)
63 print(is.factor(levelfact))
64 print(levels(levelfact))
65 print(class(levelfact))
66

```

## OUTPUT-

```

Console Terminal x Jobs x
~/
aueu_packages
> library (Rweka)
> rating <-1:4
> friends <-c ('Kuhu','Saumya','Rohan','Gomu')
> school<-c('St.X','SVM','DAV','Lotus')
> avg_school_hour <-c(21,18,17,10)
> super_friends<-data.frame(rating,friends,school,avg_schoo
l_hour,stringAsFactors=FALSE)
> print(super_friends)
  rating friends school avg_school_hour stringAsFactors
1      1   Kuhu   St.X             21          FALSE
2      2  Saumya   SVM             18          FALSE
3      3   Rohan   DAV             17          FALSE
4      4    Gomu  Lotus             10          FALSE
> print(class(super_friends))
[1] "data.frame"
> write.arff(super_friends,file="super_friends.arff")
> library("Rweka")
> library("Rweka")
> N=read.arff("super_friends.arff")
> N=read.arff("super_friends.arff")
> # Print Data
> print(N)

```

```

Console Terminal x Jobs x
~/
> print(N)
  rating friends school avg_school_hour stringAsFactors
1      1   Kuhu   St.X             21          FALSE
2      2  Saumya   SVM             18          FALSE
3      3   Rohan   DAV             17          FALSE
4      4    Gomu  Lotus             10          FALSE
> # Cat is used so that the newline characters are treated
as string and not vectors
> cat("\n\n\n")

> # Printing first two rows from the data set
> print(head(N,2))
  rating friends school avg_school_hour stringAsFactors
1      1   Kuhu   St.X             21          FALSE
2      2  Saumya   SVM             18          FALSE
> # To find the dimensions of data set
> dim(N)
[1] 4 5
> # To find the names of variables (column names) in data s
et
> names(N)

```

```

Console Terminal x Jobs x
~/
[1] rating friends school
[4] "avg_school_hour" "stringAsFactors"
> # Show all the friends
> N["friends"]
friends
1 Kuhu
2 Saumya
3 Rohan
4 Gomu
> # Create a vector.
> x <- c(12,7,3,4.2,18,2,54,-21,8,-5)
> # Find Mean.
> r<- mean(x)
> print(r)
[1] 8.22
> median(sort(x))
[1] 5.6
> # To generate a summary of data
> summary(N)
      rating      friends      school
Min.   :1.00  Length:4      Length:4
1st Qu.:1.75  Class :character  Class :character
Median :2.50  Mode  :character  Mode  :character

Console Terminal x Jobs x
~/
Max.   :4.00
avg_school_hour stringAsFactors
Min.   :10.00  Mode :logical
1st Qu.:15.25  FALSE:4
Median :17.50
Mean   :16.50
3rd Qu.:18.75
Max.   :21.00
> level<-c('Topper','Medium','Average','Below Average')
> print(class(level))
[1] "character"
> print(is.factor(level))
[1] FALSE
> levelfact<-factor(level)
> print(is.factor(levelfact))
[1] TRUE
> print(levels(levelfact))
[1] "Average"      "Below Average" "Medium"
[4] "Topper"
> print(class(levelfact))
[1] "factor"
>

```