



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment 1.2

Student Name: Kanishk Soni

Branch: BE-CSE

Semester: 6th

Subject Name: Data Mining Lab

UID: 20BCS9398

Section/Group: 20BCS-DM_708B

Date of Performance: 06-03-2023

Subject Code: 20CSP-376

1. Aim:

To perform the statistical analysis of data.

2. Objective:

To perform functions based on analysing the data by using multiple inbuilt functions of RStudio.

3. Code and Output:

PROGRAM

```
#reading the arff file
```

```
r = read.arff("D:\\College\\Sem6\\DM_Lab\\Ex1_df.arff")
```

```
print(r)
```

```
print(head(r,4))
```

```
print(tail(r,4))
```

```
#check the attributes
```

```
names(r)
```

```
# to check the dimensions of the data set
```

```
dim(r)
```

```
# to check the max and min of the marks
```

```
max(r["marks"])
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
min(r["marks"])
```

```
# finding mean and median of the marks
```

```
mean(marks)
```

```
median(sort(marks))
```

```
# standard deviation of marks
```

```
sd(marks)
```

```
# sort the marks in ascending order
```

```
sort(marks)
```

```
#summary of data
```

```
summary(r)
```

Console:

```
> #reading the arff file
```

```
> r = read.arff("D:\\College\\Sem6\\DM_Lab\\Ex1_df.arff") Error in  
read.arff("D:\\College\\Sem6\\DM_Lab\\Ex1_df.arff") :  
could not find function "read.arff"
```

```
> print(r)
```

	roll	studs	marks	status_factor
1	1	AAA	44	P
2	2	BBB	49	P
3	3	CCC	37	F
4	4	DDD	41	P
5	5	EEE	29	F
6	6	FFF	32	F
7	7	GGG	45	P

```
>
```

```
> print(head(r,4))
```

	roll	studs	marks	status_factor
1	1	AAA	44	P
2	2	BBB	49	P
3	3	CCC	37	F
4	4	DDD	41	P

```
> print(tail(r,4))
```



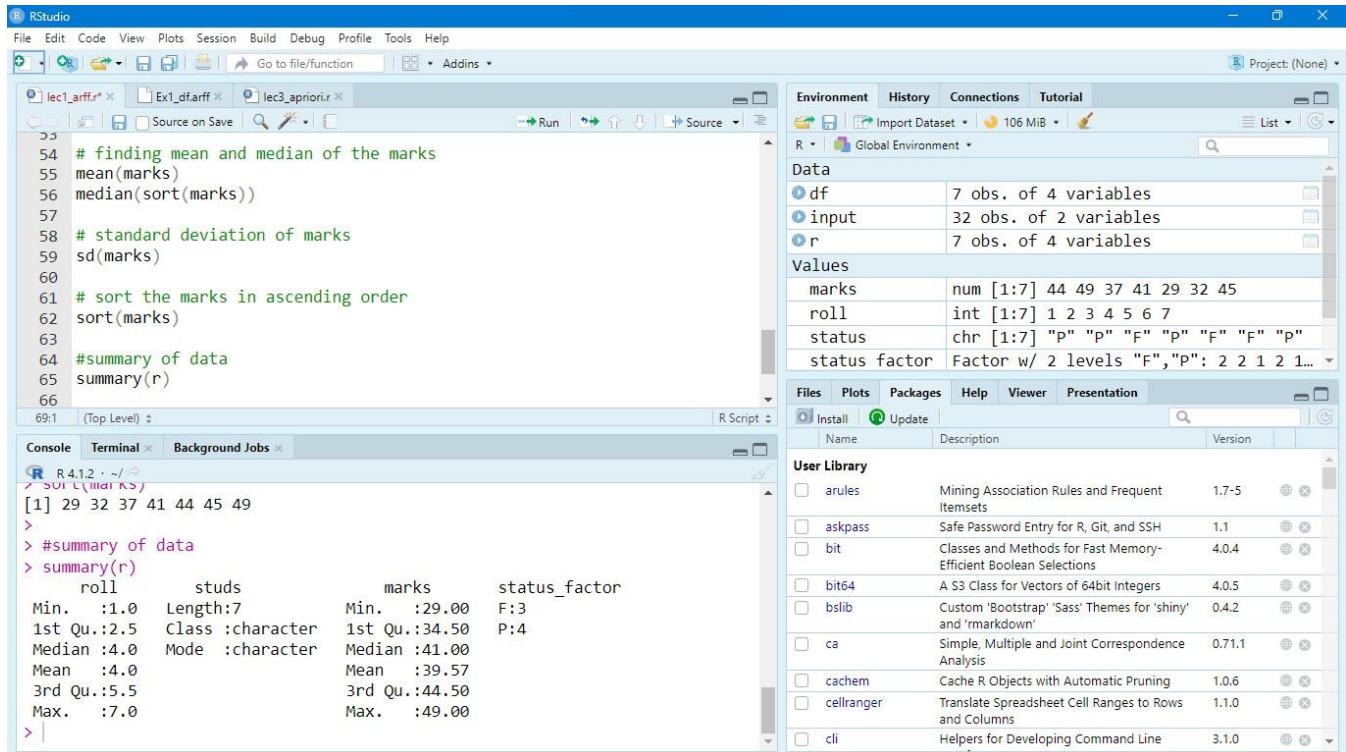
DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
roll studs marks status_factor
4 4 DDD 41 P
5 5 EEE 29 F
6 6 FFF 32 F
7 7 GGG 45 P
>
> #check the attributes
> names(r)
[1] "roll" "studs" "marks" "status_factor"
>
> # to check the dimensions of the data set
> dim(r)
[1] 7 4
> # to check the max and min of the marks
> max(r["marks"])
[1] 49
> min(r["marks"])
[1] 29
>
> # finding mean and median of the marks
> mean(marks)
[1] 39.57143
> median(sort(marks))
[1] 41
>
> # standard deviation of marks
> sd(marks)
[1]
7.253899
>
> # sort the marks in ascending order
> sort(marks)
[1] 29 32 37 41 44 45 49
>
> #summary of data
> summary(r)
      roll      studs      marks      status_factor
Min.   :1.0   Length:7   Min.   :29.00   F:3
1st Qu.:2.5   Class  :character 1st Qu.:34.50   P:4
Median :4.0   Mode   :character  Median :41.00
Mean   :4.0                      Mean   :39.57
3rd Qu.:5.5                      3rd Qu.:44.50
Max.   :7.0                      Max.   :49.00
```

OUTPUT:

RStudio:



The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains R code for finding mean, median, standard deviation, and sorting marks.
- Console:** Displays the output of the executed code, including the sorted marks and a summary of the data.
- Environment:** Shows the objects created in the global environment, including 'df', 'input', and 'r'.
- Files:** Lists the files in the current project.
- Packages:** Lists the installed and available packages.

Source Editor Code:

```
54 # finding mean and median of the marks
55 mean(marks)
56 median(sort(marks))
57
58 # standard deviation of marks
59 sd(marks)
60
61 # sort the marks in ascending order
62 sort(marks)
63
64 #summary of data
65 summary(r)
66
```

Console Output:

```
R 4.1.2 ~ /
> sort(marks)
[1] 29 32 37 41 44 45 49
>
> #summary of data
> summary(r)
      roll      studs      marks      status_factor
Min.   :1.0   Length:7   Min.   :29.00   F:3
1st Qu.:2.5   Class :character 1st Qu.:34.50   P:4
Median :4.0   Mode  :character   Median :41.00
Mean   :4.0                                Mean   :39.57
3rd Qu.:5.5                                3rd Qu.:44.50
Max.   :7.0                                Max.   :49.00
>
```

Environment:

Object	Class	Attributes
df	data.frame	7 obs. of 4 variables
input	data.frame	32 obs. of 2 variables
r	data.frame	7 obs. of 4 variables

Values:

Variable	Class	Values
marks	num [1:7]	44 49 37 41 29 32 45
roll	int [1:7]	1 2 3 4 5 6 7
status	chr [1:7]	"p" "p" "F" "p" "F" "F" "p"
status_factor	Factor w/ 2 levels	"F", "p": 2 2 1 2 1...

4. Learning Outcomes:

- Statistical analysis means investigating trends, patterns, and relationships using quantitative data.
- This type of data analysis is usually used to derive patterns, trends or other meaningful conclusions.