

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

“Experiment 1.2”

Student Name: SUMIT KUMAR

UID: 20BCS8226

Branch: CSE

Section/Group: DM-720 A

Semester: 6th

Date of Performance: 06-03-23

Subject Name: Data Mining Lab

Subject Code: 20CSP-376

Aim:

To perform the statistical analysis of data.

Objective:

- The objective is to find mean, median, mode of any provided dataset.
- To understand implementation of names and dim functions.
- To know how to display all available datasets.

Code:

```
#to display all available datasets  
library(help="datasets")
```

```
#select any dataset  
iris
```

DEPARTMENT OF

COMPUTER SCIENCE & ENGINEERING

```
#to show all attributes of the dataset  
names(iris)
```

```
#to display number of rows and columns  
dim(iris)
```

```
#to calculate the mean  
mean = mean(AirPassengers)  
print(mean)
```

```
#to calculate median  
median=median(AirPassengers)  
print(median)
```

```
#to calculate mode  
mode= function(iris){  
  return(names(sort(-table(iris$Sepal.Length)))[1])  
}  
mode(iris)
```

Output:

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains two lines of R code:

```
1 #to display all available datasets
2 library(help="datasets")
3
```
- Console:** Shows the output of the executed code:

```
> library(help="datasets")
> iris
      Sepal.Length Sepal.Width Petal.Length Petal.Width
1          5.1         3.5         1.4         0.2
2          4.9         3.0         1.4         0.2
3          4.7         3.2         1.3         0.2
4          4.6         3.1         1.5         0.2
5          5.0         3.6         1.4         0.2
6          5.4         3.9         1.7         0.4
7          4.6         3.4         1.4         0.3
8          5.0         3.4         1.5         0.2
9          4.4         2.9         1.4         0.2
10         4.9         3.1         1.5         0.1
11         5.4         3.7         1.5         0.2
12         4.8         3.4         1.6         0.2
13         4.8         3.0         1.4         0.1
14         4.3         3.0         1.1         0.1
```
- Environment:** Shows the Global Environment with variables v1 (num [1:3] 5 5 5), x (4), and y (4).
- Files:** Shows the file explorer with various files and folders, including .RData, .Rhistory, AD_Port, Arduino, Avatar, Custom Office Templates, and desktop.ini.

The screenshot shows the RStudio environment with the following components:

- Source Editor:** Contains R code for data analysis:

```
> #to show all attributes of the dataset
> names(iris)
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"
[5] "Species"
> #to display number of rows and columns
> dim(iris)
[1] 150 5
> #to calculate the mean
> mean = mean(AirPassengers)
> print(mean)
[1] 280.2986
> #to calculate median
> median=median(AirPassengers)
> print(median)
[1] 265.5
> #to calculate mode
> mode= function(iris){
+   return(names(sort(-table(iris$Sepal.Length)))[1])
+ }
> mode(iris)
[1] "5"
```
- Console:** Shows the output of the executed code, including the names of the dataset attributes, dimensions, mean, median, and mode.
- Environment:** Shows the Global Environment with variables v1 (num [1:3] 5 5 5), x (4), and y (4).
- Files:** Shows the file explorer with various files and folders, including .RData, .Rhistory, AD_Port, Arduino, Avatar, Custom Office Templates, and desktop.ini.