



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

Experiment-1.2

Student Name: Nabha Varshney

UID: 20BCS4995

Branch: CSE

Section/Group: 20BCS-DM-704 (A)

Semester: 6th

Date of Performance: 17th Feb 2023

Subject Name: Data Mining

Subject Code: 20CSP- 351

Aim - To perform the statistical analysis of data.

Objective-

- ♦ To perform statistical analysis like mean, median, standard deviation, min, max etc. after reading the data frame from stored file in system.

Script and Output-

```
library("RWeka")
```

```
N = read.arff("Data.arff")  
print(N)
```

```
print(head(N,2))  
print(tail(N,2))
```

```
dim(N)
```

```
names(N)
```

```
N["name"]
```

```
max(N["height"])
```

```
min(height)
```

```
sum(height)
```



DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
mean(height)
```

```
median(sort(height))
```

```
sd(height)
```

```
summary(N)
```

```
attributes(N)
```

```
typeof("name")
```

RStudio console output:

```
R 4.2.2 ~ ~/R> library("rweka")
R 4.2.2 ~ ~/R> N = read.arff("data.arff")
R 4.2.2 ~ ~/R> print(N)
  rating name height city avg_sleep_hour
1     1 Nabha   180  Delhi           20
2     2 Nikhil  175 Garhwal          21
3     3 Yash   177  Jaipur           23
4     4 DJ    172  Guhawati          24
5     5 Arsh   169  Chandigarh        25

R 4.2.2 ~ ~/R> print(head(N,2))
  rating name height city avg_sleep_hour
1     1 Nabha   180  Delhi           20
2     2 Nikhil  175 Garhwal          21

R 4.2.2 ~ ~/R> print(tail(N,2))
  rating name height city avg_sleep_hour
4     4 DJ    172  Guhawati          24
5     5 Arsh   169  Chandigarh        25

R 4.2.2 ~ ~/R> dim(N)
[1] 5 5

R 4.2.2 ~ ~/R> names(N)
[1] "rating"      "name"        "height"      "city"
[5] "avg_sleep_hour"

R 4.2.2 ~ ~/R> N["name"]
  name
1 Nabha
2 Nikhil
3 Yash
4 DJ
5 Arsh

R 4.2.2 ~ ~/R> max(N["height"])
[1] 180

R 4.2.2 ~ ~/R> min(height)
[1] 169

R 4.2.2 ~ ~/R> sum(height)
[1] 873

R 4.2.2 ~ ~/R> mean(height)
[1] 174.6

R 4.2.2 ~ ~/R> median(sort(height))
[1] 175

R 4.2.2 ~ ~/R> sd(height)
[1] 4.27785

R 4.2.2 ~ ~/R> summary(N)
  rating name height city
Min.   :1 Length:5 Min.   :169.0 Length:5
```

Environment pane:

Object	Class	Attributes
Data	data.frame	5 obs. of 5 variables
Datal	data.frame	4 obs. of 4 variables
N	data.frame	5 obs. of 5 variables

Values pane:

Variable	Class	Values
avg_sleep_hour	num	[1:5] 20 21 23 24 25
city	chr	[1:5] "Delhi" "Garhwal" "Jaipur" "Guhawati" "Chandiga"
country	chr	[1:4] "india" "india" "india" "india"

Files pane:

Name	Description	Version
base	The R Base Package	4.2.2
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-28
class	Functions for Classification	7.3-20
cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.1.4
codetools	Code Analysis Tools for R	0.2-18
compiler	The R Compiler Package	4.2.2
datasets	The R Datasets Package	4.2.2
foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'SyStat', 'Weka', 'dBase',...	0.8-83
graphics	The R Graphics Package	4.2.2
grDevices	The R Graphics Devices and Support for Colours and Fonts	4.2.2
grid	The Grid Graphics Package	4.2.2
kernelSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-20
lattice	Trellis Graphics for R	0.20-45
MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-58.1

RStudio console output:

```
R 4.2.2 ~ ~/R> max(N["height"])
[1] 180

R 4.2.2 ~ ~/R> min(height)
[1] 169

R 4.2.2 ~ ~/R> sum(height)
[1] 873

R 4.2.2 ~ ~/R> mean(height)
[1] 174.6

R 4.2.2 ~ ~/R> median(sort(height))
[1] 175

R 4.2.2 ~ ~/R> sd(height)
[1] 4.27785

R 4.2.2 ~ ~/R> summary(N)
  rating name height city
Min.   :1 Length:5 Min.   :169.0 Length:5
1st Qu.:2 Class :character 1st Qu.:172.0 Class :character
Median :3 Mode  :character Median :175.0 Mode  :character
Mean   :3          Mean   :174.6
3rd Qu.:4          3rd Qu.:177.0
Max.   :5          Max.   :180.0

R 4.2.2 ~ ~/R> attributes(N)
$names
[1] "rating"      "name"        "height"      "city"
[5] "avg_sleep_hour"

$class
[1] "data.frame"

$row.names
[1] 1 2 3 4 5

R 4.2.2 ~ ~/R> typeof("name")
[1] "character"
R 4.2.2 ~ ~/R> |
```

Environment pane:

Object	Class	Attributes
Data	data.frame	5 obs. of 5 variables
Datal	data.frame	4 obs. of 4 variables
N	data.frame	5 obs. of 5 variables

Values pane:

Variable	Class	Values
avg_sleep_hour	num	[1:5] 20 21 23 24 25
city	chr	[1:5] "Delhi" "Garhwal" "Jaipur" "Guhawati" "Chandiga"
country	chr	[1:4] "india" "india" "india" "india"

Files pane:

Name	Description	Version
base	The R Base Package	4.2.2
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-28
class	Functions for Classification	7.3-20
cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.1.4
codetools	Code Analysis Tools for R	0.2-18
compiler	The R Compiler Package	4.2.2
datasets	The R Datasets Package	4.2.2
foreign	Read Data Stored by 'Minitab', 'S', 'SAS', 'SPSS', 'Stata', 'SyStat', 'Weka', 'dBase',...	0.8-83
graphics	The R Graphics Package	4.2.2
grDevices	The R Graphics Devices and Support for Colours and Fonts	4.2.2
grid	The Grid Graphics Package	4.2.2
kernelSmooth	Functions for Kernel Smoothing Supporting Wand & Jones (1995)	2.23-20
lattice	Trellis Graphics for R	0.20-45
MASS	Support Functions and Datasets for Venables and Ripley's MASS	7.3-58.1