



## Experiment 1.2

**Student Name:** Praduman Kumar

**UID:** 20BCS9446

**Branch:** BE-CSE

**Section/Group:** 20BCS-DM-714/A

**Semester:** 6

**Subject Code:** 20CSP-376

**Subject Name:** Data Mining Lab

**Date of Performance:** 09-03-2023

**1. Aim/Overview of the practical:** To perform the statistical analysis of data.

**2. Tools used:** RStudio and RWeka

**3. Code:**

**Creating arrf file.R:**

```
library(RWeka)
setwd("C:\\Users\\hp\\Documents\\DATA MINING
CODES\\EXPERIMENT 2")
getwd()
rating <- 1:4
animal <- c('koala', 'hedgehog', 'sloth', 'panda')
country <- c('Australia', 'Italy', 'Peru', 'China')
avg_sleep_hours <- c(21, 18, 17, 10)
super_sleepers <- data.frame(rating, animal, country, avg_sleep_hours,
stringAsFactors=FALSE)
print(super_sleepers)
print(class(super_sleepers))
print(str(super_sleepers))
write.arff(super_sleepers, file="super_sleepers.arff")
```

**Experiment2.R:**

```
library("RWeka")
N = read.arff("super_sleepers.arff")
print(N)
```



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
cat("\n\n\n")
print(head(N,2))
dim(N)
names(N)
N["animal"]
N["avg_sleep_hours"]
max(N["avg_sleep_hours"])
min(avg_sleep_hours)
sum(avg_sleep_hours)
mean(avg_sleep_hours)
median(sort(avg_sleep_hours))
sd(avg_sleep_hours)
summary(N)
```

#### 4. Output:

RStudio:



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source

Console Terminal Background Jobs

R 4.2.2 · ~/DATA MINING CODES/EXPERIMENT 2/
> library(Rweka)
> setwd("C:\\Users\\hp\\Documents\\DATA MINING CODES\\EXPERIMENT 2")
> getwd()
[1] "C:/Users/hp/Documents/DATA MINING CODES/EXPERIMENT 2"
> rating <- 1:4
> animal <- c('koala', 'hedgehog', 'sloth', 'panda')
> country <- c('Australia', 'Italy', 'Peru', 'China')
> avg_sleep_hours <- c(21, 18, 17, 10)
> super_sleepers <- data.frame(rating, animal, country, avg_sleep_hours, stringAsFactors=FALSE)
> print(super_sleepers)
  rating animal country avg_sleep_hours stringAsFactors
1      1  koala Australia             21             FALSE
2      2 hedgehog  Italy              18             FALSE
3      3  sloth   Peru              17             FALSE
4      4  panda   China              10             FALSE
> print(class(super_sleepers))
[1] "data.frame"
> print(str(super_sleepers))
'data.frame':   4 obs. of  5 variables:
 $ rating      : int  1 2 3 4
 $ animal      : chr  "koala" "hedgehog" "sloth" "panda"
 $ country     : chr  "Australia" "Italy" "Peru" "China"
 $ avg_sleep_hours: num  21 18 17 10
 $ stringAsFactors: logi  FALSE FALSE FALSE FALSE
NULL
> write.arff(super_sleepers, file="super_sleepers.arff")
> library("Rweka")
> N = read.arff("super_sleepers.arff")
> print(N)
  rating animal country avg_sleep_hours stringAsFactors
1      1  koala Australia             21             FALSE
2      2 hedgehog  Italy              18             FALSE
3      3  sloth   Peru              17             FALSE
4      4  panda   China              10             FALSE
> cat("\n\n")

> print(head(N,2))
  rating animal country avg_sleep_hours stringAsFactors
1      1  koala Australia             21             FALSE
2      2 hedgehog  Italy              18             FALSE
> dim(N)
[1] 4 5
> names(N)
[1] "rating"          "animal"          "country"          "avg_sleep_hours" "stringAsFactors"
> N["animal"]
  animal
1  koala
2 hedgehog
3  sloth
4  panda
> N["avg_sleep_hours"]
  avg_sleep_hours
1              21
2              18
3              17
4              10
> max(N["avg_sleep_hours"])
[1] 21
> min(avg_sleep_hours)
[1] 10
> sum(avg_sleep_hours)
[1] 66
> mean(avg_sleep_hours)
[1] 16.5
> median(sort(avg_sleep_hours))
[1] 17.5
> sd(avg_sleep_hours)
[1] 4.654747
> summary(N)
  rating      animal      country      avg_sleep_hours      stringAsFactors
Min.   :1.00   Length:4   Length:4   Min.   :10.00   Mode :logical
1st Qu.:1.75   Class :character   Class :character   1st Qu.:15.25   FALSE:4
Median :2.50   Mode :character   Mode :character   Median :17.50
Mean   :2.50
3rd Qu.:3.25
Max.   :4.00
```








# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

## Files:

> Documents > DATA MINING CODES > EXPERIMENT 2

Name	Date modified	Type	Size
 .RData	09-03-2023 19:15	RDATA File	3 KB
 .Rhistory	09-03-2023 19:15	RHISTORY File	4 KB
 Creating_arff_file.R	09-03-2023 19:03	R File	1 KB
 Experiment2.R	09-03-2023 19:00	R File	1 KB
 super_sleepers	09-03-2023 19:06	ARFF Data File	1 KB

## Weka:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **None** Apply Stop

Current relation: Relation: R\_data\_frame Instances: 4 Attributes: 5 Sum of weights: 4

Attributes: All None Invert Pattern

No.	Name
1	<input type="checkbox"/> rating
2	<input type="checkbox"/> animal
3	<input type="checkbox"/> country
4	<input checked="" type="checkbox"/> avg_sleep_hours
5	<input type="checkbox"/> stringAsFactors

Remove

Selected attribute: Name: avg\_sleep\_hours Missing: 0 (0%) Distinct: 4 Type: Numeric Unique: 4 (100%)

Statistic	Value
Minimum	10
Maximum	21
Mean	16.5
StdDev	4.655

Class: stringAsFactors (Nom) Visualize All

4

10 15.5 21

Status: OK Log x 0



# DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Discover. Learn. Empower.

## 5. Observation:

- Learnt how to use R and create a file in Rstudio.
- Learnt how to import files in weka.
- Learnt about different methods like setwd() and getwd()
- Learnt about statistical methods and their usage.