Experiment1.2

Student Name: Aayush Gurung   **UID:20BCS5323**
Branch: CSE   **Section/Group:DM_607(A)**
Semester: 6   **Date of Performance:16-02-2023**
Subject Name: DM LAB   **Subject Code: 20CSP-376**

1. **Aim:**

   To perform the statistical analysis of data.

2. **Code:**

```
# Perform Statistical analysis of data

library("RWeka")

N = read.arff("super_sleepers.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))

# 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 animals
N["animal"]

# Show Average sleep hours
N["avg_sleep_hours"]

# Show max of average sleep hours
max(N["avg_sleep_hours"])

# Show min of average sleep hours
min(avg_sleep_hours)
```

```r
# Sum of average sleep hours
sum(avg_sleep_hours)

# Mean of average sleep hours
mean(avg_sleep_hours)

# Median of average sleep hours
median(sort(avg_sleep_hours))

# Standard Deviation of average sleep hours
sd(avg_sleep_hours)

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

3. Output:

```
Source

Console   Terminal ×   Background Jobs ×
R  R 4.2.2 · C:/Users/aayush/Desktop/
>
> library("RWeka")
> N = read.arff("super_sleepers.arff")
> # Print Data
> 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 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   animal   country avg_sleep_hours stringAsFactors
1      1    koala Australia              21           FALSE
2      2 hedgehog     Italy              18           FALSE
> # To find the dimensions of data set
> dim(N)
[1] 4 5
> # To find the names of variables (Column names) in data set
> names(N)
[1] "rating"          "animal"          "country"          "avg_sleep_hours" "stringAsFactors"
> # Show all the animals
> N["animal"]
    animal
1    koala
2 hedgehog
3    sloth
4    panda
> # Show Average sleep hours
> N["avg_sleep_hours"]
  avg_sleep_hours
1              21
2              18
3              17
4              10
> # Show max of average sleep hours
```

Console **Terminal** × **Background Jobs** ×

R 4.2.2 · C:/Users/aayush/Desktop/

```
1      koala
2 hedgehog
3      sloth
4      panda
> # Show Average sleep hours
> N["avg_sleep_hours"]
  avg_sleep_hours
1              21
2              18
3              17
4              10
> # Show max of average sleep hours
> max(N["avg_sleep_hours"])
[1] 21
> # Show min of average sleep hours
> min(avg_sleep_hours)
[1] 10
> # Sum of average sleep hours
> sum(avg_sleep_hours)
[1] 66
> # Mean of average sleep hours
> mean(avg_sleep_hours)
[1] 16.5
> # Median of average sleep hours
> median(sort(avg_sleep_hours))
[1] 17.5
> # Standard Deviation of average sleep hours
> sd(avg_sleep_hours)
[1] 4.654747
> # To generate a summary of data
> 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                                          Mean   :16.50
 3rd Qu.:3.25                                          3rd Qu.:18.75
 Max.   :4.00                                          Max.   :21.00
>
> |
```