



**DEPARTMENT OF**

**COMPUTER SCIENCE & ENGINEERING**

Discover. Learn. Empower.

### Experiment1.2

**Student Name: Arif khan**

**Branch: BE CSE**

**Semester: 6<sup>TH</sup>**

**UID: 20BCS4947**

**Section/Group: 20BCS\_DM\_607\_B**

**Subject Name: DM LAB**

- 1. Aim: Statistical analysis of data.**
- 2. Objective: To perform statistical analysis of data.**
- 3. Script and Output:**

```
library(RWeka) #setting  
the working directory  
setwd("C:\\Users\\dell\\OneDrive\\Desktop\\DataMining")  
#checking the working  
directory getwd()
```

```
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/ ↗  
> library(RWeka)  
> setwd("C:\\Users\\dell\\OneDrive\\Desktop\\DataMining")  
> getwd()  
[1] "C:/Users/dell/OneDrive/Desktop/DataMining"  
> |
```

```
#reading arff file  
N = read.arff("diabetes  
.arff") print(N)
```

```
6:1 (Top Level) R Script
Console Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
98 1 71 48 18 76 20.4 0.323 22 tested_negative
99 6 93 50 30 64 28.7 0.356 23 tested_negative
100 1 122 90 51 220 49.7 0.325 31 tested_positive
101 1 163 72 0 0 39.0 1.222 33 tested_positive
102 1 151 60 0 0 26.1 0.179 22 tested_negative
103 0 125 96 0 0 22.5 0.262 21 tested_negative
104 1 81 72 18 40 26.6 0.283 24 tested_negative
105 2 85 65 0 0 39.6 0.930 27 tested_negative
106 1 126 56 29 152 28.7 0.801 21 tested_negative
107 1 96 122 0 0 22.4 0.207 27 tested_negative
108 4 144 58 28 140 29.5 0.287 37 tested_negative
109 3 83 58 31 18 34.3 0.336 25 tested_negative
110 0 95 85 25 36 37.4 0.247 24 tested_positive
111 3 171 72 33 135 33.3 0.199 24 tested_positive
[ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> |
```

**#printing first and  
last 2 rows head(N,2)  
tail(N,2)**

```
10:1 (Top Level) R Script
Console Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
106 1 126 56 29 152 28.7 0.801 21 tested_negative
107 1 96 122 0 0 22.4 0.207 27 tested_negative
108 4 144 58 28 140 29.5 0.287 37 tested_negative
109 3 83 58 31 18 34.3 0.336 25 tested_negative
110 0 95 85 25 36 37.4 0.247 24 tested_positive
111 3 171 72 33 135 33.3 0.199 24 tested_positive
[ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> head(N,2)
  preg plas pres skin insu mass pedi age      class
1     6  148   72   35    0 33.6 0.627  50 tested_positive
2     1   85   66   29    0 26.6 0.351  31 tested_negative
> tail(N,2)
  preg plas pres skin insu mass pedi age      class
767    1  126   60    0    0 30.1 0.349  47 tested_positive
768    1   93   70   31    0 30.4 0.315  23 tested_negative
> |
```

**#printing  
dimension dim(N)**

```

14:1 (Top Level) R Script
Console Background Jobs x
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
108 4 144 58 28 140 29.5 0.287 37 tested_negative
109 3 83 58 31 18 34.3 0.336 25 tested_negative
110 0 95 85 25 36 37.4 0.247 24 tested_positive
111 3 171 72 33 135 33.3 0.199 24 tested_positive
[ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> head(N,2)
  preg plas pres skin insu mass pedi age      class
1    6  148  72  35    0 33.6 0.627  50 tested_positive
2    1   85  66  29    0 26.6 0.351  31 tested_negative
> tail(N,2)
  preg plas pres skin insu mass pedi age      class
767    1  126  60    0    0 30.1 0.349  47 tested_positive
768    1   93  70  31    0 30.4 0.315  23 tested_negative
> dim(N)
[1] 768  9
>

```

## #printing column

### names names(N)

```

17:1 (Top Level) R Script
Console Background Jobs x
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
110 0 95 85 25 36 37.4 0.247 24 tested_positive
111 3 171 72 33 135 33.3 0.199 24 tested_positive
[ reached 'max' / getOption("max.print") -- omitted 657 rows ]
> head(N,2)
  preg plas pres skin insu mass pedi age      class
1    6  148  72  35    0 33.6 0.627  50 tested_positive
2    1   85  66  29    0 26.6 0.351  31 tested_negative
> tail(N,2)
  preg plas pres skin insu mass pedi age      class
767    1  126  60    0    0 30.1 0.349  47 tested_positive
768    1   93  70  31    0 30.4 0.315  23 tested_negative
> dim(N)
[1] 768  9
> names(N)
[1] "preg" "plas" "pres" "skin" "insu" "mass" "pedi" "age" "class"
>

```

## #printing maximum and minimum

### age

max(N["age"])

min(N["age"])

```

24:10 (Top Level) R Script
Console Background Jobs x
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
> preg plas pres skin insu mass pedi age class
1 6 148 72 35 0 33.6 0.627 50 tested_positive
2 1 85 66 29 0 26.6 0.351 31 tested_negative
> tail(N,2)
preg plas pres skin insu mass pedi age class
767 1 126 60 0 0 30.1 0.349 47 tested_positive
768 1 93 70 31 0 30.4 0.315 23 tested_negative
> dim(N)
[1] 768 9
> names(N)
[1] "preg" "plas" "pres" "skin" "insu" "mass" "pedi" "age" "class"
> max(N["age"])
[1] 81
> min(N["age"])
[1] 21
>

```

**#mean age**

**mean(N\$age)**

```

24:1 (Top Level) R Script
Console Background Jobs x
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
> dim(N)
[1] 768 9
>
> #printing column names
> names(N)
[1] "preg" "plas" "pres" "skin" "insu" "mass" "pedi" "age" "class"
>
> #printing maximum and minimum age
> max(N["age"])
[1] 81
> min(N["age"])
[1] 21
>
> #mean age
> mean(N$age)
[1] 33.24089
>

```

**#median age**

**median(sort(N\$age))**

```
27:1 (Top Level) R Script
Console Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
>
> #printing column names
> names(N)
[1] "preg" "plas" "pres" "skin" "insu" "mass" "pedi" "age" "class"
>
> #printing maximum and minimum age
> max(N["age"])
[1] 81
> min(N["age"])
[1] 21
> #mean age
> mean(N$age)
[1] 33.24089
> median(sort(N$age))
[1] 29
> |
```

## #standard deviation

sd(N\$age)

```
31:10 (Top Level) R Script
Console Background Jobs
R 4.2.2 · C:/Users/dell/OneDrive/Desktop/DataMining/
> names(N)
[1] "preg" "plas" "pres" "skin" "insu" "mass" "pedi" "age" "class"
>
> #printing maximum and minimum age
> max(N["age"])
[1] 81
> min(N["age"])
[1] 21
> #mean age
> mean(N$age)
[1] 33.24089
> median(sort(N$age))
[1] 29
> sd(N$age)
[1] 11.76023
> |
```

## #summary

summary(N)

```

33:1 (Top Level) R Script
Console Background Jobs x
R 4.2.2 C:/Users/dell/OneDrive/Desktop/DataMining/
> summary(N)
      preg      plas      pres      skin      insu
Min.   : 0.000  Min.   : 0.0  Min.   : 0.00  Min.   : 0.00  Min.   : 0.0
1st Qu.: 1.000  1st Qu.: 99.0  1st Qu.: 62.00  1st Qu.: 0.00  1st Qu.: 0.0
Median : 3.000  Median :117.0  Median : 72.00  Median :23.00  Median : 30.5
Mean   : 3.845  Mean   :120.9  Mean   : 69.11  Mean   :20.54  Mean   : 79.8
3rd Qu.: 6.000  3rd Qu.:140.2  3rd Qu.: 80.00  3rd Qu.:32.00  3rd Qu.:127.2
Max.   :17.000  Max.   :199.0  Max.   :122.00  Max.   :99.00  Max.   :846.0

      mass      pedi      age      class
Min.   : 0.00  Min.   :0.0780  Min.   :21.00  tested_negative:500
1st Qu.:27.30  1st Qu.:0.2437  1st Qu.:24.00  tested_positive:268
Median :32.00  Median :0.3725  Median :29.00
Mean   :31.99  Mean   :0.4719  Mean   :33.24
3rd Qu.:36.60  3rd Qu.:0.6262  3rd Qu.:41.00
Max.   :67.10  Max.   :2.4200  Max.   :81.00
>

```

## 5. Learning Outcomes:

- ☐ Creating data frames.
- ☐ Reading an .arff file.
- ☐ Worked on large dataset and analyse it.
- ☐ Printing dataset and performing statistical analysis .
- ☐ Finding mean, median, standard deviation etc of the data frame

1.