Experiment No: 2 Date:

Descriptive Statistics Using R Programming

Aim: Implement Descriptive Statistics Using R Programming

Theory:

Descriptive statistics is a branch of statistics aiming at summarizing, describing and presenting a series of values or a dataset. Descriptive statistics is often the first step and an important part in any statistical analysis. It allows checking the quality of the data and it helps to "understand" the data by having a clear overview of it. If well presented, descriptive statistics is already a good starting point for further analyses.

In this program the dataset Iris is used. This dataset is imported default in R. The dataset is loaded using the following command

```
dat <- iris # load the iris dataset and renamed it dat
```

A preview of this dataset and its structure can be viewed using the command:

```
head(dat) # first 6 observations
```

Structure of the dataset can be also viewed:

```
str(dat) # structure of dataset
```

The dataset contains 150 observations and 5 variables, representing the length and width of the sepal and petal and the species of 150 flowers. Length and width of the sepal and petal are numeric variables and the species is a factor with 3 levels.

Minimum and maximum can be found thanks to the *min()* and *max()* functions. Alternatively the *range()* function can also be used.

The mean can be computed with the *mean()* function.

The median can be computed thanks to the *median()* function or with the *quantile()* function.

As the median, the first and third quartiles can be computed thanks to the *quantile()* function and by setting the second argument to 0.25 or 0.75.

The standard deviation and the variance is computed with the *sd()* and *var()* functions.

It is also possible to compute the minimum, 1st quartile, median, mean, 3rd quartile and the maximum for all numeric variables of a dataset at once using *summary()*.

Program Printout:

```
library(datasets)

data <- iris


repeat {


        cat("\n\n1.  Head\n2.  Structure\n3.  Max\n4.  Min\n5.  Mean\n6.  Median  1st  and  3rd  quartile\n7.  Standard  deviation  and  variance\n8.  Summary\n9.  Exit\n\nEnter your choice: ")

    choice <- as.integer(readline())


    if(!(choice %in% 1:9))

    {

        cat("\n\nInvalid choice!")

        next

    }


    if(choice %in% 3:7){

        name = readline("Enter the name of the column: ")

        if(!(name %in% colnames(data)))

        {

            print("Column not found!")

            next

        }

    }
```

```
    switch(
        choice,
        print(head(data)),
        print(str(data)),
        print(max(data[[name]])),
        print(min(data[[name]])),
        print(mean(data[[name]])),
        {
        print(paste("Median: ",median(data[[name]])))
           print(paste("1st Quartile: ",quantile(data[[name]],
0.25)))
           print(paste("3rd Quartile: ",quantile(data[[name]],
0.75)))
        },
        {
        print(paste("Standard Deviation: ",sd(data[[name]])))
        print(paste("Variance: ",var(data[[name]])))
        },
        print(summary(data)),
        {break},
    )

}
```

Program Output:

```
> source("exp2.r")


 1. Head
 2. Structure
 3. Max
 4. Min
 5. Mean
 6. Median 1st and 3rd quartile
 7. Standard deviation and variance
 8. Summary
 9. Exit

 Enter your choice: 1
    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
 1          5.1         3.5          1.4         0.2  setosa
 2          4.9         3.0          1.4         0.2  setosa
 3          4.7         3.2          1.3         0.2  setosa
 4          4.6         3.1          1.5         0.2  setosa
 5          5.0         3.6          1.4         0.2  setosa
 6          5.4         3.9          1.7         0.4  setosa
```

```
1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

Enter your choice: 2
 'data.frame':  150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1
1 1 ...
NULL
```

```
Enter your choice: 3
 Enter the name of the column: Petal.Length
[1] 6.9


1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

Enter your choice: 4
 Enter the name of the column: Petal
[1] "Column not found!"


1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

4
 Enter the name of the column: Petal.Length
[1] 1
```

```
1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

Enter your choice: 5
 Enter the name of the column: Petal.Length
[1] 3.758


1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

Enter your choice: 6
 Enter the name of the column: Petal.Width
[1] "Median:  1.3"
[1] "1st Quartile:  0.3"
[1] "3rd Quartile:  1.8"
```

```
1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

Enter your choice: 7
 Enter the name of the column: Sepal.Width
[1] "Standard Deviation:  0.435866284936698"
[1] "Variance:  0.189979418344519"


1. Head
2. Structure
3. Max
4. Min
5. Mean
6. Median 1st and 3rd quartile
7. Standard deviation and variance
8. Summary
9. Exit

Enter your choice: 8
   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
 Median :5.800   Median :3.000   Median :4.350   Median :1.300
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
       Species
 setosa    :50
 versicolor:50
 virginica :50
```

```
Enter your choice: 10


Invalid choice!
```

```
9. Exit

Enter your choice: 9
>
```

Conclusion: The experiment successfully implemented Descriptive Statistics in R using the Iris dataset.