



Finolex Academy of Management and Technology, Ratnagiri

Department of Information Technology

Subject:	R Programming Lab. (ITL804)		
Class:	BE IT / Semester – VIII (Rev-2016) / Academic year: 2019-20		
Name of Student:	Kazi Jawwad A Rahim		
Roll No:	28	Date of performance (DOP):	21/02/2020
Assignment/Experiment No:	04	Date of checking (DOC):	23/03/2020
Title: Exploratory data analysis such as- Range, summary, mean, variance, median, standard deviation, histogram, boxplot, scatterplot			
Marks:	100	Teacher's Signature:	

1. Aim: To understand the exploratory data analysis and the methods required to do it in R.

2. Prerequisites:

1. Basics of R programming, various data structures, functions etc.

3. Hardware Requirements:

1. PC with minimum 2GB RAM

4. Software Requirements:

1. Windows / Linux OS.
2. R version 3.6 or higher

5. Learning Objectives:

1. To understand decision and loop control instructions.
2. To understand function definition and calling to it.

6. Learning Objectives Applicable: LO 3. LO 4

7. Program Outcomes Applicable: PO 2, PO 3

8. Program Education Objectives Applicable: PEO 2, PEO 3



Theory:

Data Analysis:

1) Range:

In statistics, the range of a set of data is the difference between the largest and smallest values.

2) Summary:

In statistics, summary of a data set gives the values of minimum, 1st quartile (25%), median (2nd quartile - 50%), mean (average), 3rd quartile and maximum.

3) Mean:

In statistics, mean of a data set represents the average value of the data set.

4) Mode:

In statistics, mode of data set is the value which occurs most number of times in the set.

5) Median:

In statistics, median of data set is the value which is at the middle of arranged or sorted data set.

6) Variance:

In statistics, variance is the expectation of the squared deviation of a random variable from its mean.

D



7) Standard Deviation:

In statistics, standard deviation is a measure of the amount of variation or dispersion of a set of values.

8) Histogram:

In statistics, a histogram is a display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size.

9) Boxplot:

In statistics, a boxplot is a method for graphically depicting groups of numerical data through their quartiles.

10) Scatter plot:

In statistics, a scatter plot is a group used to determine whether there is a relationship between paired data.

Range:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

range(a)

OUTPUT:

[1] 1 9

Summary:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

summary(a)

OUTPUT:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	5.000	6.500	6.417	8.250	9.000

Mean:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

mean(a)

OUTPUT:

[1] 6.416667

Mode:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

table(a)

OUTPUT:

a

1 4 5 6 7 8 9

1 1 2 2 1 2 3

=> Mode=9

Median:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

median(a)

OUTPUT:

[1] 6.5

Variance:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

var(a)

OUTPUT:

[1] 5.901515

Standard Deviation:

a=c(1,5,7,8,9,6,4,9,5,8,9,6)

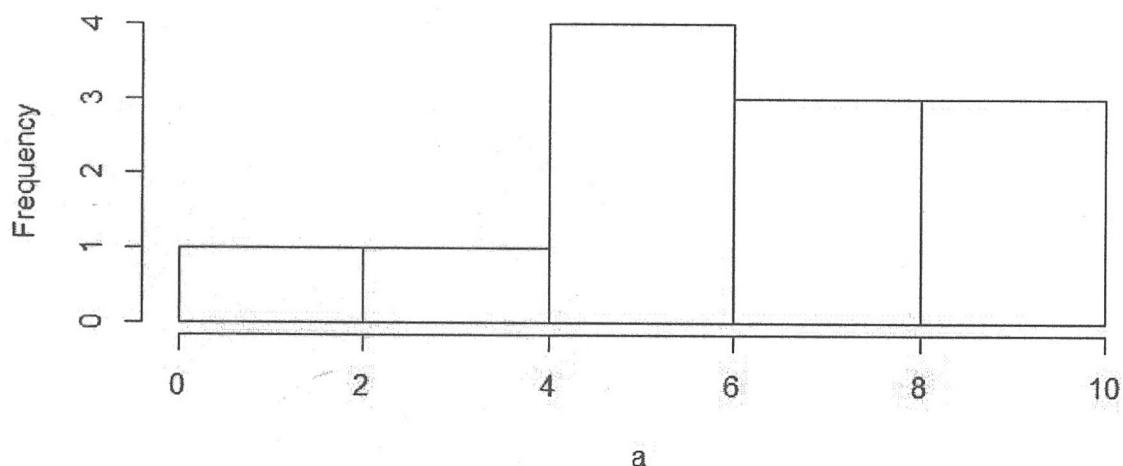
sqrt(var(a))

OUTPUT:

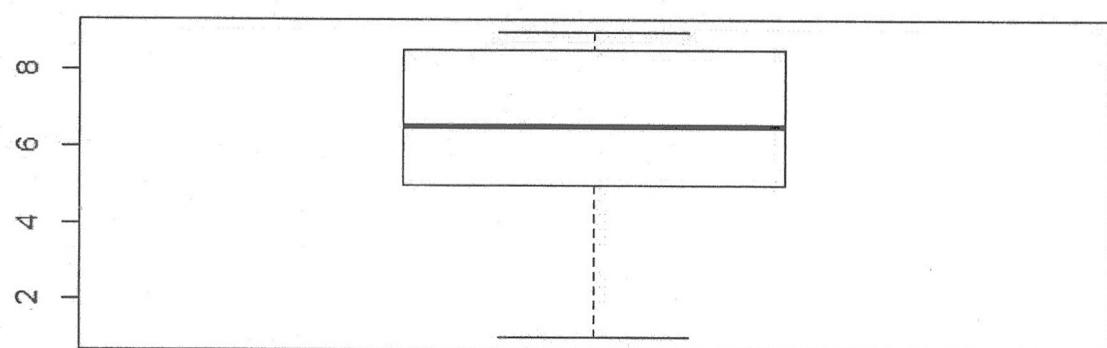
[1] 2.429303

Histogram:

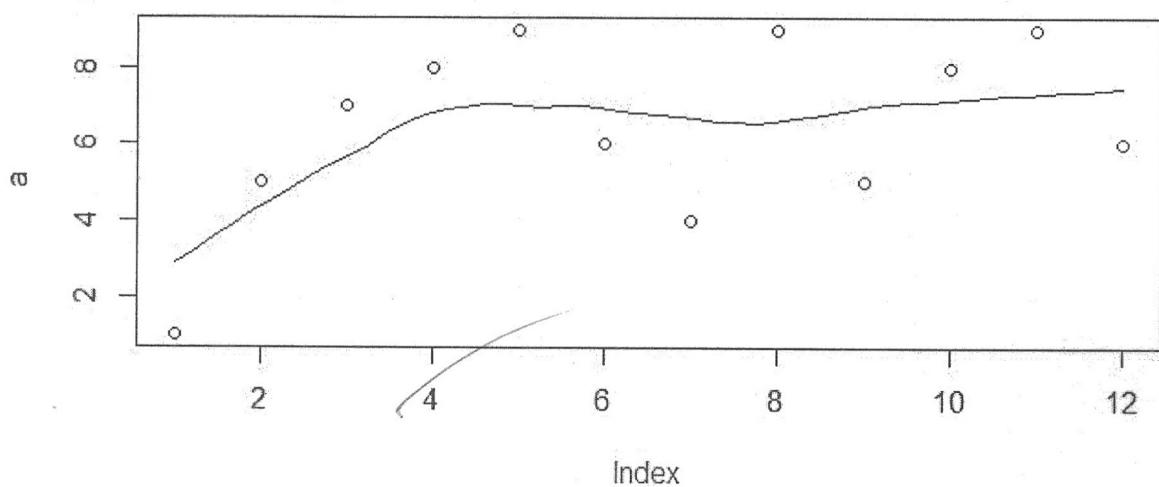
Histogram of a



Boxplot:



Scatterplot:



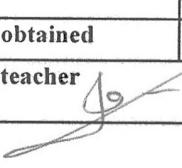
Learning Outcomes Achieved:

1. We understood decision and loop control instructions.
2. We understood the function definition and its calling.

Conclusion:

We have successfully demonstrated the data analysis such as- Range, summary, mean, variance, median, standard deviation, histogram, boxplot, scatterplot.

13. Experiment/Assignment Evaluation

Experiment/Assignment Evaluation:			
Sr. No.	Parameters	Marks obtained	Out of
1	Technical Understanding (Assessment may be done based on Q & A <u>or</u> any other relevant method.) Teacher should mention the other method used -	86	6
2	Neatness/presentation	62	2
3	Punctuality	61	2
Date of performance (DOP)	21/02/2020	Total marks obtained	10
Date of checking (DOC)	06/03/2020	Signature of teacher	

References:

1. URL: <https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf> (Online Resources)
2. R Cookbook Paperback – 2011 by Teetor Paul O Reilly Publications
3. Beginning R: The Statistical Programming Language by Dr. Mark Gardener, Wiley Publications
4. R Programming For Dummies by Joris Meys Andrie de Vries, Wiley Publications

Viva Questions

1. What is exploratory data analysis ?
2. What is summary of the data ?
3. What is importance of median of the data collection ?
4. What is histogram? Why is it important in data?
5. What information does the box plot provides?
6. List various R library functions used in exploratory data analysis.