EX.NO: 11	STATISTICAL MEASURE USING R PROGRAMMING
DATE: 31-10-2022	

Aim:

To perform some basic statistical measure using R.

Description:

1) Minimum

min function returns the minimum value of a vector or column.

```
> min(data$Sepal.Length)
[1] 4.3
>
```

2) Maximum

max function returns the maximum value of a vector or column.

```
> max(data$Sepal.Length)
[1] 7.9
>
```

3) Mean

The **mean** of an observation variable is a numerical measure of the central location of the data values. It is the sum of its data values divided by data count.

```
> mean(data$Sepal.Length)
[1] 5.843333
>
```

4) Median

The **median** of an observation variable is the value at the middle when the data is sorted in ascending order. It is an ordinal measure of the central location of the data values.

```
> median(data$Sepal.Length)
[1] 5.8
>
```

5) Mode

The mode value is the value that appears the most number of times. R does not have a function to calculate the mode.

We can easily find mode using the functions table() and sort() :

```
> tab <- table(data$Sepal.Length)
> sort(tab, decreasing = TRUE)

  5 5.1 6.3 5.7 6.7 5.5 5.8 6.4 4.9 5.4 5.6   6 6.1 4.8 6.5 4.6 5.2 6.2 6.9 7.7 4.4
 10   9   9   8   8   7   7   7   6   6   6   6   6   5   5   4   4   4   4   4   3
5.9 6.8 7.2 4.7 6.6 4.3 4.5 5.3   7 7.1 7.3 7.4 7.6 7.9
  3   3   3   2   2   1   1   1   1   1   1   1   1   1
>
```

6) Range

The **range** of an observation variable is the difference of its largest and smallest data values. It is a measure of how far apart the entire data spreads in value.

Range = Largest value – Smallest value

```
> range(data$Sepal.Length)
[1] 4.3 7.9
>
```

7) Quartiles

Quartiles divide the entire set into four equal parts. So, there are three quartiles, first, second and third represented by Q_1, Q_2 and Q_3, respectively. Q_2 is nothing but the median

```
> quantile(data$Sepal.Length, 0.25)
25%
5.1
>
> quantile(data$Sepal.Length, 0.75)
75%
6.4
>
```

8) Interquartile Range

The **interquartile range** of an observation variable is the difference of its upper and lower quartiles. It is a measure of how far apart the middle portion of data spreads in value.

IQR = Upper Quartile – Lower Quartile

```
> IQR(data$Sepal.Length)
[1] 1.3
>
```

9) Standard Deviation

The **standard deviation** of an observation variable is the square root of its variance.

```
> sd(data$Sepal.Length)
[1] 0.8280661
>
```

10) Variance

The variance is a numerical measure of how the data values is dispersed around the mean.

```
> var(data$Sepal.Length)
[1] 0.6856935
>
```

11) Summary

summary is a generic function used to produce result summaries of the results of various model fitting functions.

```
> summary(data)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
 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
>
```

12) Covariance

Covariance is the measure of the relation between two variables of a dataset. That is, it depicts the way two variables are related to each other.

```
> x <- c(1, 3, 5, 10)
> y <- c(2, 4, 6, 20)
> cov(x, y)
[1] 30.66667
>
```

13) Correlation

Correlation on a statistical basis is the method of finding the relationship between the variables in terms of the movement of the data. That is, it helps us analyze the effect of changes made in one variable over the other variable of the dataset.

```
> x <- c(1, 3, 5, 10)
> y <- c(2, 4, 6, 20)
> cor(x, y, method = "pearson")
[1] 0.9724702
>
```

14)Percentile

Percentile is a comparison score between a particular score and the scores of the rest of a group. It shows the percentage of scores that a particular score surpassed.

```
> x <- c(2,13,5,36,12,50)
> quantile(x,probs=c(0.25,0.5,0.75))
  25%   50%   75%
 6.75 12.50 30.25
>
```

Result:

Thus, successfully completed statistical measures using r programming.