

Semester : VISubject : DAV

Academic Year: 2023-2024

DESCRIPTIVE STATISTICS:

In descriptive analysis, we are describing our data with the help of excel files, graphs, tables etc and present it in a meaningful way so that it can be easily understood. Most of the times it is performed on small datasets and this analysis help us a lot to predict some future trends based on the current findings. Some measures that are used to describe a dataset are

- \* measures of central tendency.
- \* measures of variability.

Measures of Central Tendency:-

- \* It represents the whole set of data by a single value.
- \* It gives us the location of central points.
- \* There are 3 main measures of central tendency:  
(i) Mean (ii) Median (iii) Mode.

Examples: Let us consider BOD dataset in R.

BOD.

This will display the dataset BOD. This dataset has 2 Column Time and demand. with 7 rows.

`myData = BOD` # Assigns BOD dataset to variable myData.

`mean = mean(myData$demand)`

`print(mean)`

This will calculate the mean of demand column of myData dataset and store it in variable mean. The print() method will display the calculated mean.

Semester: VISubject: DAVAcademic Year: 20 ~~23~~ 2024

```
median = median(myData$demand)
```

```
print(median)
```

This method will display the median of the demand column.

Measures of variability:-

Measure of variability is known as the spread of data or how well is our data is distributed. The most common variability measures are.

(i) Range (ii) Variance (iii) Standard Deviation.

```
max = max(myData$demand)
```

```
print(max)
```

```
min = min(myData$demand)
```

```
print(min)
```

```
range = max - min.
```

```
print(range).
```

The max() and min() methods displays the maximum and minimum values of demand column. The range is calculated by finding the difference between max and the min.

```
variance = var(myData$demand)
```

```
print(variance)
```

```
sd = sd(myData$demand)
```

```
print(sd)
```

The var() method will calculate the variance of demand column. The sd() method calculates the standard deviation in R.



Semester : VISubject : DAVAcademic Year: 20 23 20 24`quartiles = quantile(BOD$demand)``print(quartiles)`

The `quantile()` method calculates and displays the 0%, 25%, 50%, 75%, 100% percentiles of the demand column in myData dataset.

`IQR = IQR(BOD$demand)``print(IQR)`

The `IQR()` method displays the InterQuartile Range of demand column.

`summary(BOD$demand)`

The `summary` method calculates and display the summary of the demand column. The output is displayed in the following way:

Min	1 <sup>st</sup> Qu.	Median	Mean	3 <sup>rd</sup> Qu.	Max.
8.80	11.62	15.80	14.83	18.25	19.80

It calculates and gives the above value. Descriptive Analysis displays the statistics of the given dataset. Statistics is all about drawing conclusions from data.