Amit Rawat

MCA - 'A' 1st Sem 1st year

21711018 - Student ID

Class Roll no - 42

AR

R programming End Sem Practical

Vinod Singh Rawat (Father Name)

Amit Rawat Student id - 21711018 Class Roll no - 42

MCA - A

R Programming

3. Analyzing CSV File

Getting and setting the working directory

```
# Get and Print current working directory.
print(getwd())
# Set current working directory.
setwd("/web/com")
# Get and print current working directory.
print(getwd())
```

Input as CSV File - input.csv

Read a CSV File

```
data <- read.csv("input.csv")
print(data)
```

Analyzing the CSV File

```
data <- read.csv("input.csv")
print(is.data.frame(data))
print(ncol(data))
print(nrow(data))
```

create a data frame.

```
data <- read.csv("input.csv")
```

Get the max salary from data frame.

```
sal <- max(data$salary)
```

Amit Rawat Student id - 21711018 Class Rollno-42

MCA - A

```
print(sal)
retval <- subset( data, salary == max( salary ))
print( retval )

# write filtered data into new file.
write.csv (retval, "output.csv")
newdata <- read.csv("output.csv")
print (newdata)
```

Amit Rawat Student id - 21711018 Class Roll no-42

MCA-A

R programming

4.

Descriptive statistics are used to describe the characterstics or features of a dataset. The term 'descriptive statistics' can be used to describe both individual quantative observations as well as the overall process of obtaining insights from these data.

Mean : The average value of all the data points.

Median : The central or middle value in the dataset.

Mode : The value that appears most often in the dataset.

Standard deviation : High standard deviation suggests that the values are more broadly spread out.

Minimum and maximum values : These are the highest and lowest values in a dataset or quantile.

Range : This measures the size of the distribution of values.

Amit Rawat Student id - 21711018 Class Roll no-42

MCA-A

Inferential Statistics

Inferential statistics focus on making generalizations about a larger population based on a representative sample of that population.

Because inferential statistics focuses on making prediction (rather than stating facts) its results are usually in the form of a probability.

Confidence intervals are used to estimate certain parameters for a measurement of a population based on sample data.

Rather than providing a single mean value, the confidance interval provides a range of values. This is often given as a percentage.

Regression and correlation analysis are both techniques used for observing how two or more sets of variables relate to one another. correlation analysies, measures the degree of association between two or more datasets.