ITA0443 STATISTICS WITH R PROGRAMING

NAME: N M S N V ARUN

REGISTER:192110529

UNIVARIATE ANALYSIS IN R – MEASURES OF CENTRAL TENDENCY

Exercise:

I. ARITHMETIC MEAN

a) Write suitable R code to compute the average of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

programm:

```
x<-(12,7,3,4.2,18,2,54,-21,8,-5)
Error: unexpected ',' in "x<-(12,"
> x<-c(12,7,3,4,2,18,2,54,-21,8,-5)
> y<-mean(x)
> print(y)
```

output:

```
[1] 7.636364
```

b) Compute the mean after applying the trim option and removing 3 values from each

end.

programm:

```
x<-c(12,7,3,4,2,18,2,54,-21,8,-5)
> y<-mean(x,trim=3)
> print(y)
[1] 4
> y<-mean(x,trim=0.3)
```

```
> print(y)
```

output:

```
[1] 4.8
```

c) Compute the mean of the following vector .

(12,7,3,4.2,18,2,54,-21,8,-5,NA)

#If there are missing values, then the mean function returns NA.

Find mean dropping NA values.

#To drop the missing values from the calculation use na.rm = TRUE

programm:

```
x<-c(12,7,3,4,2,18,2,54,-21,8,-5,NA)
> y<-mean(x)
> print(y)
[1] NA
> y<-mean(x,na.rm=TRUE)
> print(y)
```

output:

```
[1] 7.636364
```

II.MEDIAN

Write suitable R code to compute the median of the following values.

12,7,3,4.2,18,2,54,-21,8,-5

programm:

```
x<-c(12,7,3,4.2,18,2,54,-21,8,-5)
> y<-median(x)
```

> print(y)

output:

[1] 5.6

III. MODE

Calculate the mode for the following numeric as well as character data set in R.

(2,1,2,3,1,2,3,4,1,5,5,3,2,3) , ("o","it","the","it","it")

programm:

```
 x <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)
> mode_numeric <- names(sort(-table(x)))[1]
> mode_numeric
y <- c("o","it","the","it","it")
> mode_character <- names(sort(-table(y)))[1]
> mode_character
```

output:

[1] "2"

[1] "it"

UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 1

Download mpg dataset which contains Fuel economy data from 1999 and 2008 for 38 popular models of car from the URL given below.

https://vincentarelbundock.github.io/Rdatasets/datasets.html

Answer the following queries

- Find the car which gives maximum city miles per gallon

  programm:

```
data(mtcars)
> boxplot(mtcars$mpg ~ mtcars$cyl, xlab="Number of Cylinders", ylab="Miles per Gallon", main="Boxplot of mpg vs cyl")
> data(mtcars)
> max_mpg <- max(mtcars$mpg)
> car_with_max_mpg <- rownames(mtcars[mtcars$mpg == max_mpg, ])
> cat("The car that gives the maximum city miles per gallon is", car_with_max_mpg)
```

output:

The car that gives the maximum city miles per gallon is Toyota Corolla

- Find the cars which gives minimum disp in compact and subcompact class

programm:

```
> data(mtcars)

> mtcars$class <- ifelse(mtcars$disp < 200, "subcompact", "compact")

> min_disp_compact <- min(mtcars[mtcars$class == "compact", ]$disp)

> min_disp_subcompact <- min(mtcars[mtcars$class == "subcompact", ]$disp)

> cars_with_min_disp_compact <- rownames(mtcars[mtcars$disp == min_disp_compact & mtcars$class == "compact", ])

> cars_with_min_disp_subcompact <- rownames(mtcars[mtcars$disp == min_disp_subcompact & mtcars$class == "subcompact", ])

> cat("The cars that give the minimum displacement in the compact class are:", cars_with_min_disp_compact, "\n")
```

The cars that give the minimum displacement in the compact class are: Valiant

```
> cat("The cars that give the minimum displacement in the subcompact class are:", cars_with_min_disp_subcompact)
```

output:

The cars that give the minimum displacement in the subcompact class are: Toyota Corolla

Exercise: 2

Use the same dataset as used in Exercise 1 and perform the following queries

- Find the standard deviation of city milles per gallon

programm:

```
> data(mtcars)
> sd_mpg <- sd(mtcars$mpg)
> cat("The standard deviation of city miles per gallon is", sd_mpg)
```

output:

The standard deviation of city miles per gallon is 6.026948

- Find the variance of highway milles per gallon

program:

```
> var(mpg$cty)
```

output:

[1] 18.11307

Exercise 3

Use the same dataset and perform the following queries

- Find the range of the disp in the data set mpg

programm:

> range(mpg$displ)

output:

[1] 1.6 7.0

- Find the Quartile of the disp in the data set mpg
- programm:
- > quantile(mpg$displ)
- output:

  0% 25% 50% 75% 100%
- 1.6 2.4 3.3 4.6 7.0
- Find the IQR of the disp column in the data set mpg
- programm:
- > IQR(mpg$displ)

output:

[1] 2.2

Exercise 4

#Install Library

library(e1071)

- Find the skewness of city miles per mileage in the data set mpg ?

  Use qplot function and display the graph for the city miles per mileage column

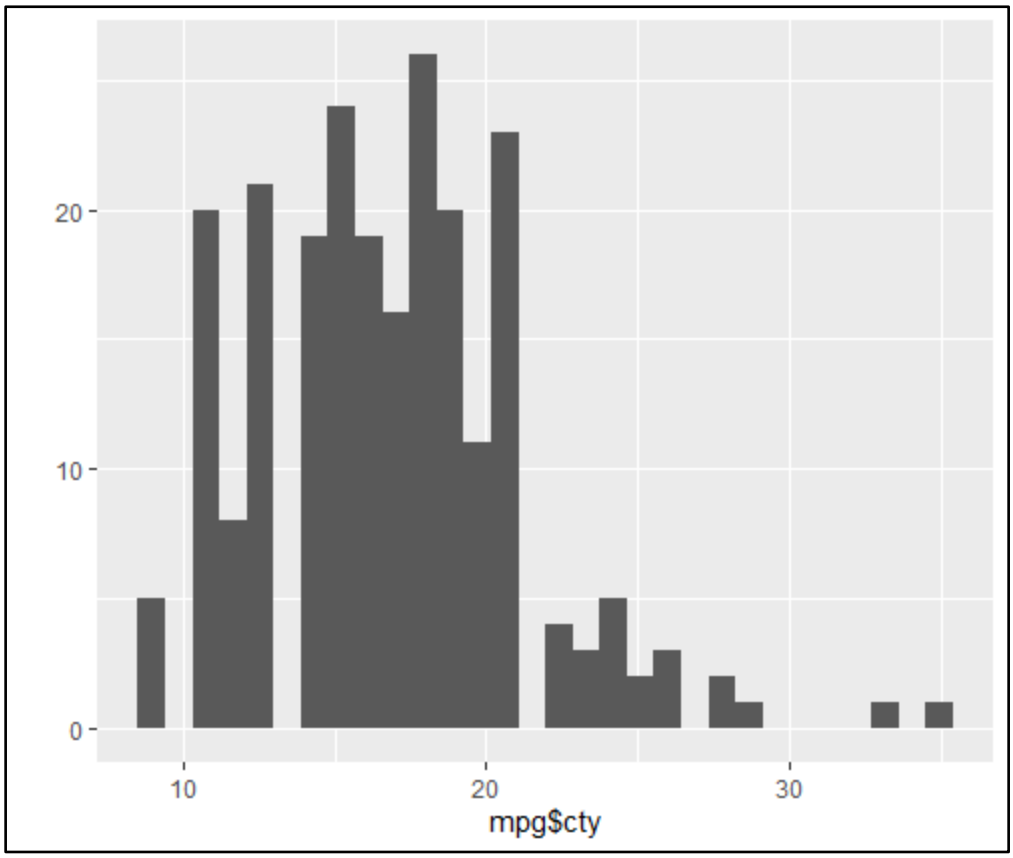
  programm:

  > library(e1071)

  > qplot(mpg$cty)

  `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

  output:



- Find the kurtosis of city miles per mileage in the data set mpg

  Use qplot function and display the graph for the city miles per mileage column

BIVARIATEANALYSIS IN R -COVARIANCE,CORRELATION,CROSSTAB

Exercise: 1

	Reference	Status	Gender	TestNewOrFollowUp
1	KRXH	Accepted	Female	Test1 New
2	KRPT	Accepted	Male	Test1 New
3	FHRA	Rejected	Male	Test2 New
4	CZKK	Accepted	Female	Test3 New
5	CQTN	Rejected	Female	Test1 New
6	PZXW	Accepted	Female	Test4 Follow-up
7	SZRZ	Rejected	Male	Test4 New
8	RMZE	Rejected	Female	Test2 New
9	STNX	Accepted	Female	Test3 New
10	TMDW	Accepted	Female	Test1 New

- Load the dataset and Create a data frame and name it as dataframe1
- Load the function for crosstab

```
xtabs(~colname , data=Data frame name )
```

Exercise: 2

- Use Two Categorical Variables and Discover the relationships within a dataset
- Next, using the xtabs() function, apply two variables from "dataframe1 ", to create a table delineating the relationship between the "Reference" category, and the "Status" category.
- Save the file in the name of dataframe2

Exercise: 3

Use the same data frame using three Categorical Variables create a Multi-Dimensional Table

Apply three variables from "dataframe1" to create a Multi-Dimensional Cross-Tabulation of "Status", "Gender", and "Test".

Exercise: 4

Row Percentages

The R package "tigerstats" is required for the next two exercises.

1) Create an xtabs() formula that cross-tabulates "Status", and "Test".

2) Enclose the xtabs() formula in the tigerstats function, "rowPerc()" to display row percentages for "Status" by "Test".

Exercise 5

Column Percentages

1) Create an xtabs() formula that cross-tabulates "Status", and "Test".

2) Enclose the xtabs() formula in the tigerstats function, "colPerc()" to display row percentages for "Status" by "Test".

Exercise 6

Covariance

- For the Dataframe1 created from exercise 2 calculate the covariance between Refrence column and Status column
- Display the covariance matrix

Exercise 7

Correlation

Find the Correlation between gender and status. what kind of correlation does exist between the two?

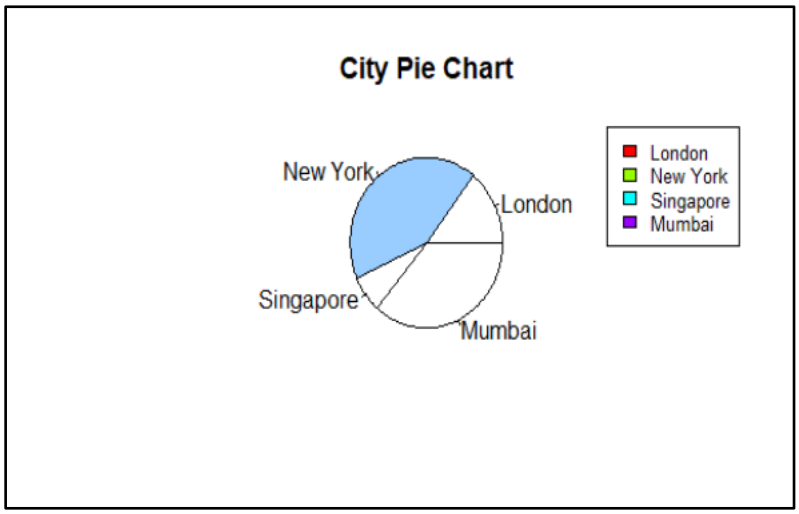
VISUALIZATION IN R

- Write a program for creating a pie-chart in R using the input vector(21,62,10,53). Provide labels for the chart as 'London', 'New York', 'Singapore', 'Mumbai'. Add a title to the chart as 'city pie-chart' and add a legend at the top right corner of the chart.

  programm:

  ```
  values <- c(21,62,10,53)
  > labels <- c("London", "New York", "Singapore", "Mumbai")
  > pie(values, labels = labels, main = "City Pie Chart")
  > legend("topright", labels, cex = 0.8, fill = rainbow(length(values)))
  >
  ```

  output:



- Create a 3D Pie Chart for the dataset "political Knowledge" with suitable labels,colours and a legend at the top right corner of the chart.

- Write a program for creating a bar chart using the vectors H=c(7,12,28,3,41) and M=c("mar", "apr", "may", "jun", "jul"). Add a title to the chart as "Revenue chart".

  programm:

  H <- c(7, 12, 28, 3, 41)

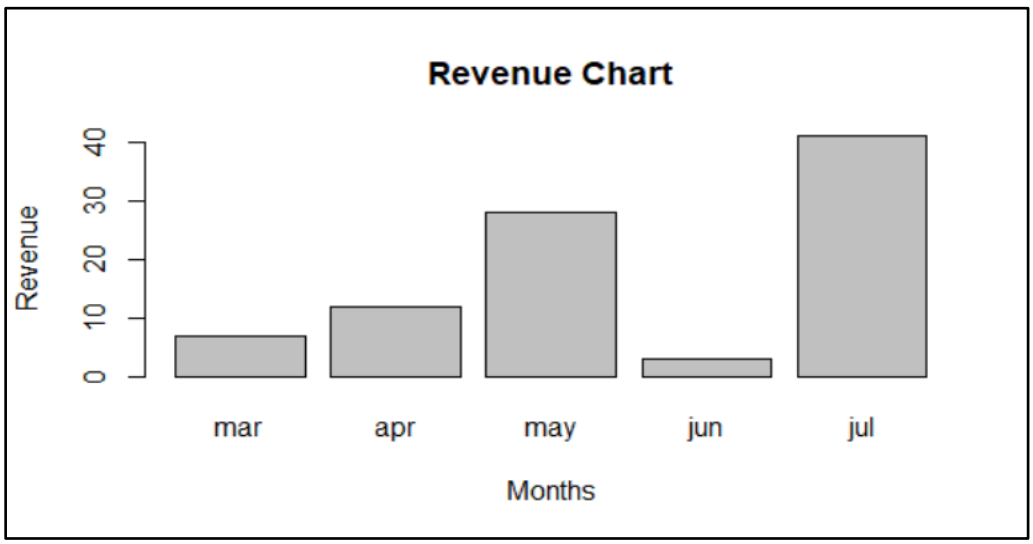
  > M <- c("mar", "apr", "may", "jun", "jul")

  >

  > barplot(H, names.arg=M, main="Revenue Chart", xlab="Months", ylab="Revenue")

  >

  output:



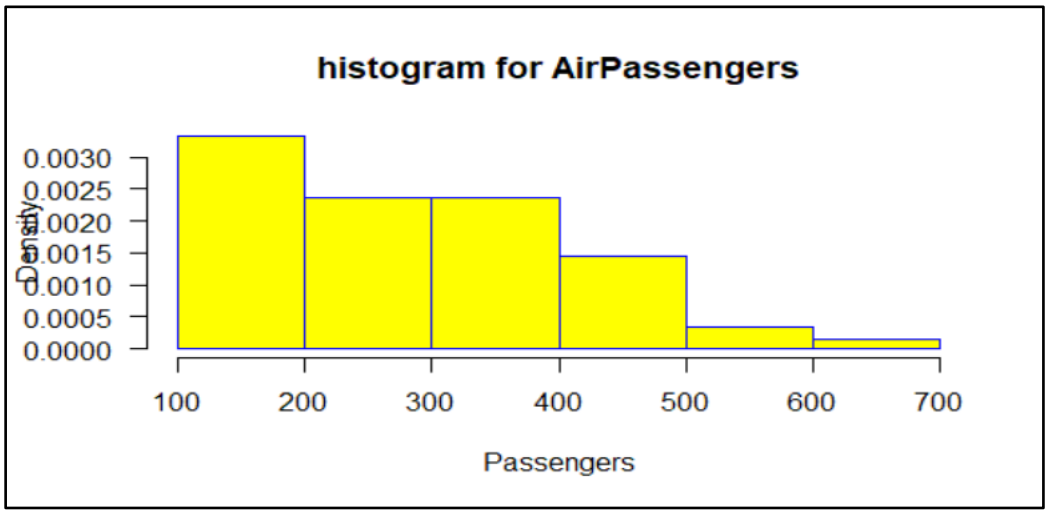
- Make a histogram for the "AirPassengers"dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 200 wide

  programm:

  hist(AirPassengers,main="histogram for AirPassengers",xlab="Passengers",border="blue",col="yellow",xlim=c(100,700),las=1,breaks=5,pro

b=TRUE)

output:



- Create a Boxplot graph for the relation between "mpg"(miles per galloon) and "cyl"(number of Cylinders) for the dataset "mtcars" available in R Environment.

program:

```
data(mtcars)

> boxplot(mtcars$mpg ~ mtcars$cyl, xlab="Number of Cylinders", ylab="Miles per Gallon", main="Boxplot of mpg vs cyl")

>
```

output: