UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY

Name : D.TEJ KIRAN

Regno : 192110625

I. ARITHMETIC MEAN

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

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

end.

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

a<-c(12,7,3,4.2,18,2,54,-21,8,-5)

average(a)

mean(a)

trimws(a,which = c("both"))

output:

```
average(a)
[1] 8.22
> mean(a)
[1] 8.22
> trimws(a,which = c("both"))
 [1] "12"  "7"   "3"   "4.2" "18"  "2"   "54"  "-21" "8"   "-5"

>
```

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

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

mean(a, na.rm = TRUE)

output:

```
1] 8.22
```

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

a<-c(12,7,3,4.2,18,2,54,-21,8,-5)

median(a)

```
> a<-c(12,7,3,4.2,18,2,54,-21,8,-5)
> median(a)
[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")

a<-c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

mode(a)

b<-c("o","it","the","it","it")

mode(b)

output:

```
> a<-c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)
> mode(a)
[1] "numeric"
> b<-c("o","it","the","it","it")
> mode(b)
[1] "character"
```

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

i) Find the car which gives maximum city miles per gallon

```
data<-mtcars
data
max_city_mpg<-max(data$mpg)
top_cars<-data[data$mpg == max_city_mpg, ]
print(top_cars)
```

output:

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2

```
AMC Javelin         15.2   8 304.0 150 3.15 3.435 17.30  0  0    3    2
Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
Pontiac Firebird    19.2   8 400.0 175 3.08 3.845 17.05  0  0    3    2
Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
Ford Pantera L      15.8   8 351.0 264 4.22 3.170 14.50  0  1    5    4
Ferrari Dino        19.7   6 145.0 175 3.62 2.770 15.50  0  1    5    6
Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
Volvo 142E          21.4   4 121.0 109 4.11 2.780 18.60  1  1    4    2
> max_city_mpg<-max(data$mpg)
>
>
> top_cars<-data[data$mpg == max_city_mpg, ]
>
>
> print(top_cars)
                mpg cyl disp hp drat    wt qsec vs am gear carb
Toyota Corolla 33.9   4 71.1 65 4.22 1.835 19.9  1  1    4    1
```

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

library(dplyr)

cars_data <- mtcars

compact_subcompact_cars <- cars_data %>%

filter(cyl %in% c(4, 5))

grouped_cars <- compact_subcompact_cars %>%

group_by(cyl)

minimum_disp <- grouped_cars %>%

summarise(min_disp = min(disp))

minimum_disp

output:

```
# A tibble: 1 × 2
    cyl min_disp
  <dbl>    <dbl>
1     4     71.1
```

Exercise: 2

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

i) Find the standard deviation of city milles per gallon

data<-mtcars

a<-sd(data$mpg)

a

output:

```
> a
[1] 6.026948
```

ii) Find the variance of highway milles per gallon

```
data<-read.csv("mpg.csv")
var(data$hwy)
```

[1] 35.45778

Exercise 3

Use the same dataset and perform the following queries

i) Find the range of the disp in the data set mpg
ii) Find the Quartile of the disp in the data set mpg
iii) Find the IQR of the disp column in the data set mpg

```
range <- max(mtcars$disp) - min(mtcars$disp)
range
quartiles <- quantile(mtcars$disp, probs = c(0.25, 0.5, 0.75))
quartiles
IQR <- quartiles[3] - quartiles[1]
IQR
```

Output:

```
range
[1] 400.9
> quartiles <- quantile(mtcars$disp, probs = c(0.25, 0.5, 0.75))
> quartiles
   25%    50%    75%
120.825 196.300 326.000
> IQR <- quartiles[3] - quartiles[1]
> IQR
   75%
205.175
>
```

Exercise 4

#Install Library

library(e1071)

a. 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

   data<-read.csv("mpg.csv")

   library(moments)

   skewness(data$cty)

   library(plotrix)

   plot(data$cty)

   output:

1] 0.7914453

b. 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

   data<-read.csv("mpg.csv")

   library(moments)

   kurtosis(data$cty)

   library(plotrix)

   plot(data$cty)

   output:

kurtosis(data$cty)

[1] 4.468651

new

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
```

Commented [nm1]:

```
8       RMZE Rejected Female Test2          New
9       STNX Accepted Female Test3          New
10      TMDW Accepted Female Test1          New
```

i) Load the dataset and Create a data frame and name it as dataframe1

ii) Load the function for crosstab

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

ref<-
c("KRXH","KRPT","FHRA","CZKK","CQTN","PZXW","SZRZ","RMZE","STNX","TMDW")

status<-
c("accepted","accepted","rejected","accepted","rejected","accepted","rejected","rejected","accepted","accepted")

gender<-
c("female","female","male","male","female","female","female","male","female","female")

test<-c("test1","test1","test2","test3","test1","test4","test4","test2","test3","test1")

neworfollowers<-c("new","new","new","new","new","follow up","new","new","new","new")

dataframe1<-data.frame(ref,status,gender,test,neworfollowers)

dataframe1

xtabs(~ref , data=dataframe1)

ouput:
ref

CQTN CZKK FHRA KRPT KRXH PZXW RMZE

1 1 1 1 1 1 1

STNX SZRZ TMDW

1 1 1

VISUALIZATION IN R

1. 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.

```
ibrary(plotrix)

a<-c(1,2,3,4,5)

labels <- c("London", "New York", "Singapore", "Mumbai")
```

```
pie(a,labels ,title("city pie chart"),col = rainbow(length(x)))
legend("topright", c("London","New York","Singapore","Mumbai"), cex = 0.8,
    fill = rainbow(length(x)))
```

2. 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.

```
data<-read.csv("Political.csv")
pie3D(data$Year,labels ="political",col = rainbow(length(x)))
```

3. 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”.

```
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",col="red")
```

4. 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

```
data<-AirPassengers
hist(data,breaks = seq(100,700,by=200),xlab = "histogram")
```

5. 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.

```
library(ggplot2)

ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
  geom_boxplot() +
  ggtitle("Boxplot of mpg by cyl") +
  xlab("Number of Cylinders") +
  ylab("Miles per Gallon")
```