**Day 3 Lab Manual PART 1**

**UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY**

**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

**Program:**

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

mean(values)

**Output:**

> mean(values)

[1] 8.22

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

end.

**Program:**

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

mean(values,trim=0.3)

**Output:**

> mean(values,trim=0.3)

[1] 5.55

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

**Program:**

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

mean(values)

mean(values,na.rm=TRUE)

**Output:**

**> mean(values)**

**[1] NA**

**> mean(values,na.rm=TRUE)**

**[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\

**Program:**

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

median(values)

**Output:**

> median(values)

[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")

**Program:**

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

mode\_num<-names(which.max(table(values)))

mode\_num

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

mode\_char<-names(which.max(table(value)))

mode\_char

**Output:**

**> mode\_num**

**[1] "2"**

**> value<-c("o","it","the","it","it")**

**> mode\_char<-names(which.max(table(value)))**

**> mode\_char**

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

**Program:**

# Download the mpg dataset

url <- "<https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv>"

mpg <- read.csv(url)

# Find the car which gives maximum city miles per gallon

max\_city\_mpg <- which.max(mpg$cty)

car\_max\_city\_mpg <- mpg$manufacturer[max\_city\_mpg]

cat("The cars which gives maximum city miles per gallon is:", car\_max\_city\_mpg)

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

compact\_class <- mpg[mpg$class == "compact", ]

min\_disp\_compact\_class <- which.min(compact\_class$displ)

car\_min\_disp\_compact\_class <- compact\_class$manufacturer[min\_disp\_compact\_class]

cat("The cars which gives minimum disp in compact class is:", car\_min\_disp\_compact\_class)

subcompact\_class <- mpg[mpg$class == "subcompact", ]

min\_disp\_subcompact\_class <- which.min(subcompact\_class$displ)

car\_min\_disp\_subcompact\_class <- subcompact\_class$manufacturer[min\_disp\_subcompact\_class]

cat("The cars which gives minimum disp in subcompact class is:", car\_min\_disp\_subcompact\_class)

**Output:**

**> cat("The car which gives maximum city miles per gallon is:", car\_max\_city\_mpg)**

**The car which gives maximum city miles per gallon is: volkswagen**

**> cat("The car which gives minimum disp in compact class is:", car\_min\_disp\_compact\_class)**

**The car which gives minimum disp in compact class is: audi**

**> cat("The car which gives minimum disp in subcompact class is:", car\_min\_disp\_subcompact\_class)**

**The car which gives minimum disp in subcompact class is: honda**

**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
* Find the variance of highway milles per gallon

**Program:**

library(ggplot2)

data("mpg")

sd\_cty<-sd(mpg$cty)

var\_hwy<-var(mpg$hwy)

cat("Standard deviation of city miles per gallon:",sd\_cty,"\n")

cat("variance of highway miles per gallon",var\_hwy)

**output:**

> cat("Standard deviation of city miles per gallon:",sd\_cty,"\n")

Standard deviation of city miles per gallon: 4.255946

> cat("variance of highway miles per gallon",var\_hwy)

variance of highway miles per gallon 35.45778

**Exercise 3**

Use the same dataset and perform the following queries

* Find the range of the disp in the data set mpg
* Find the Quartile of the disp in the data set mpg
* Find the IQR of the disp column in the data set mpg

**Program:**

library(ggplot2)

data("mpg")

disp\_range<-range(mpg$disp)

disp\_quantile<-quantile(mpg$disp)

disp\_IQR<-IQR(mpg$disp)

cat("Range of the sidp column:",disp\_range,"\n")

cat("Quartile of the disp column:",disp\_quantile,"\n")

cat("IQR of the disp column:",disp\_IQR)

**Output:**

**> cat("Range of the sidp column:",disp\_range,"\n")**

**Range of the sidp column: Inf -Inf**

**> cat("Quartile of the disp column:",disp\_quantile,"\n")**

**Quartile of the disp column: NA NA NA NA NA**

**> cat("IQR of the disp column:",disp\_IQR)**

**IQR of the disp column: NA**

**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

**program:**

library(e1071)

skew\_cty<-skewness(mpg$cty)

cat("Skewness of city miles per gallon:",skew\_cty,"\n")

**Output:**

> cat("Skewness of city miles per gallon:",skew\_cty,"\n")

Skewness of city miles per gallon: 0.7863773

* 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

**program:**

library(e1071)

kurt\_cty<-kurtosis(mpg$ctya)

cat("Kurtosis of city gallon:",kurt\_cty)

**Output:**

> cat("Kurtosis of city gallon:",kurt\_cty)

Kurtosis of city gallon: NaN