

Day 3 Lab Manual

UNIVARIATE ANALYSIS IN R - MEASURES OF CENTRAL TENDENCY

Exercise:3

NAME:KRISHVANTH KUMAR E

REGISTER NUMBER:192125063

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

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

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

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

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")

CODE

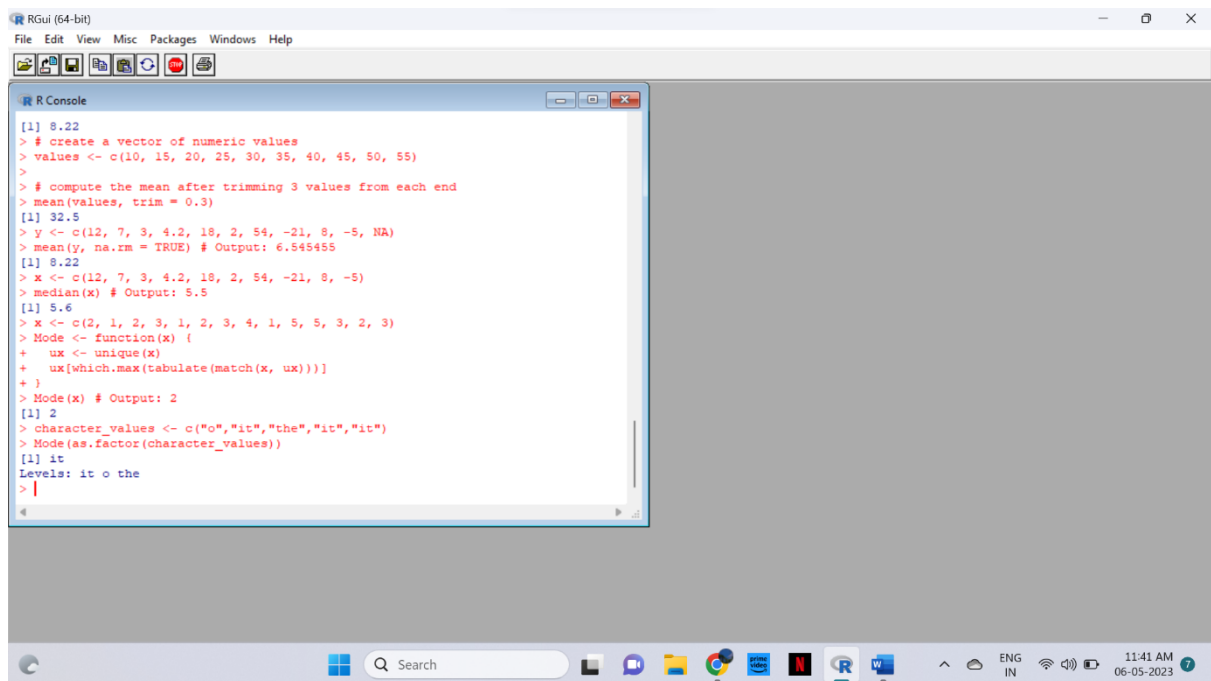
```
x <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
mean(x)
values <- c(10, 15, 20, 25, 30, 35, 40, 45, 50, 55)
mean(values, trim = 0.3)
y <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5, NA)
mean(y, na.rm = TRUE)
x <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
median(x)
x <- c(2, 1, 2, 3, 1, 2, 3, 4, 1, 5, 5, 3, 2, 3)
Mode <- function(x) {
  ux <- unique(x)
  ux[which.max(tabulate(match(x, ux)))]
}
```

```
}
```

```
Mode(x)
```

```
character_values <- c("o","it","the","it","it")
```

```
Mode(as.factor(character_values))
```



The screenshot shows the RGui (64-bit) window. The R Console contains the following code and output:

```
[1] 8.22
> # create a vector of numeric values
> values <- c(10, 15, 20, 25, 30, 35, 40, 45, 50, 55)
>
> # compute the mean after trimming 3 values from each end
> mean(values, trim = 0.3)
[1] 32.5
> y <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5, NA)
> mean(y, na.rm = TRUE) # Output: 6.545455
[1] 8.22
> x <- c(12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
> median(x) # Output: 5.5
[1] 5.6
> x <- c(2, 1, 2, 3, 1, 2, 3, 4, 1, 5, 5, 3, 2, 3)
> Mode <- function(x) {
+   ux <- unique(x)
+   ux[which.max(tabulate(match(x, ux)))]
+ }
> Mode(x) # Output: 2
[1] 2
> character_values <- c("o","it","the","it","it")
> Mode(as.factor(character_values))
[1] it
Levels: it o the
> |
```

UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION

Exercise: 4

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

CODE

```
mpg <-
read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")

mpg[which.max(mpg$cty),]

mpg[mpg$class %in% c("compact", "subcompact") & mpg$displ ==
min(mpg[mpg$class %in% c("compact", "subcompact"),]$displ),]
```

```
RGui (64-bit)
File Edit View Misc Packages Windows Help

R Console
71.100 120.825 196.300 326.000 472.000
> IQR(mtcars$displ)
[1] 205.175
>
>
> data(mpg)
Warning message:
In data(mpg) : data set 'mpg' not found
> sd(mpg$cty)
Error: object 'mpg' not found
> mpg <- read.csv("https://vincentarelbundock.github.io/Rdatasets/csv/ggplot2/mpg.csv")
> mpg[which.max(mpg$cty),]
      X manufacturer      model displ year cyl      trans drv  cty   hwy fl      class
222 222   volkswagen new beetle   1.9 1999    4 manual(m5)   f   35   44   d      compact
222 222   volkswagen new beetle   1.9 1999    4 manual(m5)   f   35   44   d      compact
> mpg[mpg$class %in% c("compact", "subcompact") & mpg$displ == min(mpg[mpg$class %in% c("compact", "subcompact"),]$displ),]
      X manufacturer      model displ year cyl      trans drv  cty   hwy fl      class
100 100     honda civic     1.6 1999    4 manual(m5)   f   28   33   r subcompact
101 101     honda civic     1.6 1999    4 auto(l4)    f   24   32   r subcompact
102 102     honda civic     1.6 1999    4 manual(m5)   f   25   32   r subcompact
103 103     honda civic     1.6 1999    4 manual(m5)   f   23   29   p subcompact
104 104     honda civic     1.6 1999    4 auto(l4)    f   24   32   r subcompact
```

Exercise: 5

Use the same dataset as used in Exercise 4 and perform the following queries

- Find the standard deviation of city miles per gallon
- Find the variance of highway miles per gallon

```
library(dplyr)
```

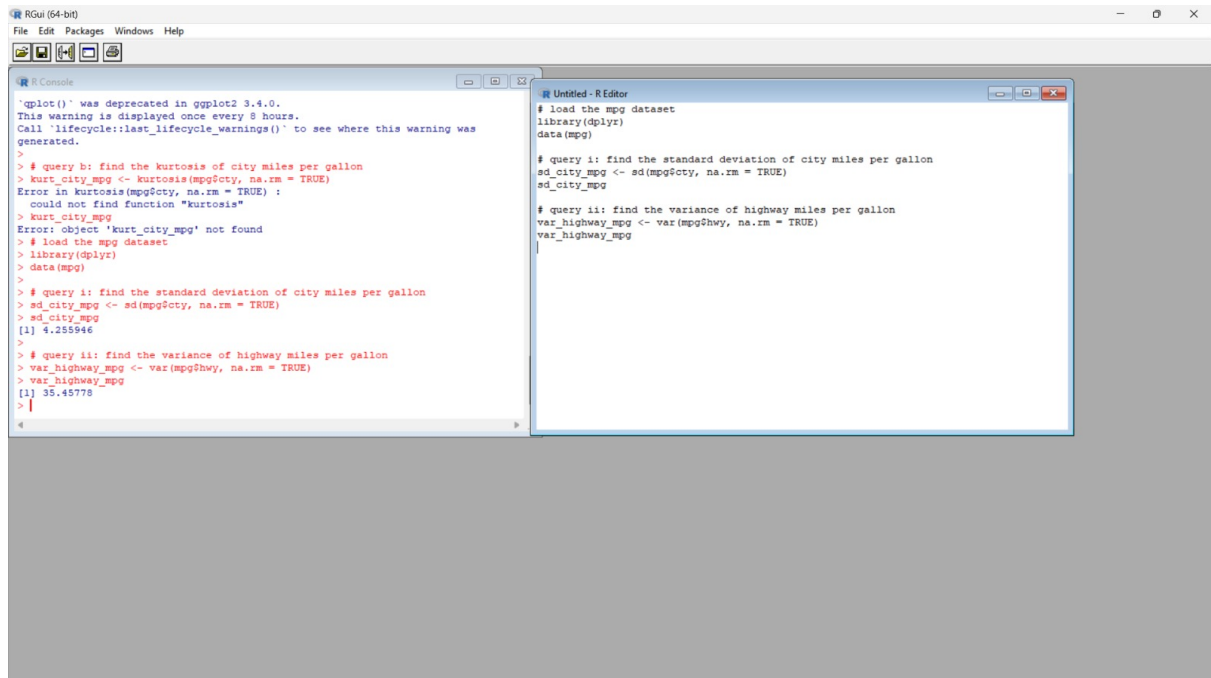
```
data(mpg)
```

```
sd_city_mpg <- sd(mpg$cty, na.rm = TRUE)
```

sd_city_mpg

```
var_highway_mpg <- var(mpg$hwyl, na.rm = TRUE)
```

var_highway_mpg



The screenshot shows an RStudio interface. The console window on the left contains the following text:

```
'qplot()' was deprecated in ggplot2 3.4.0.  
This warning is displayed once every 8 hours.  
Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was  
generated.  
>  
> # query b: find the kurtosis of city miles per gallon  
> kurt_city_mpg <- kurtosis(mpg$cty, na.rm = TRUE)  
Error in kurtosis(mpg$cty, na.rm = TRUE) :  
could not find function "kurtosis"  
> kurt_city_mpg  
Error: object 'kurt_city_mpg' not found  
> # load the mpg dataset  
> library(dplyr)  
> data(mpg)  
>  
> # query i: find the standard deviation of city miles per gallon  
> sd_city_mpg <- sd(mpg$cty, na.rm = TRUE)  
> sd_city_mpg  
[1] 4.255946  
>  
> # query ii: find the variance of highway miles per gallon  
> var_highway_mpg <- var(mpg$hwyl, na.rm = TRUE)  
> var_highway_mpg  
[1] 35.45778  
> |
```

The script editor window on the right contains the following code:

```
# load the mpg dataset  
library(dplyr)  
data(mpg)  
  
# query i: find the standard deviation of city miles per gallon  
sd_city_mpg <- sd(mpg$cty, na.rm = TRUE)  
sd_city_mpg  
  
# query ii: find the variance of highway miles per gallon  
var_highway_mpg <- var(mpg$hwyl, na.rm = TRUE)  
var_highway_mpg
```

Exercise 6

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

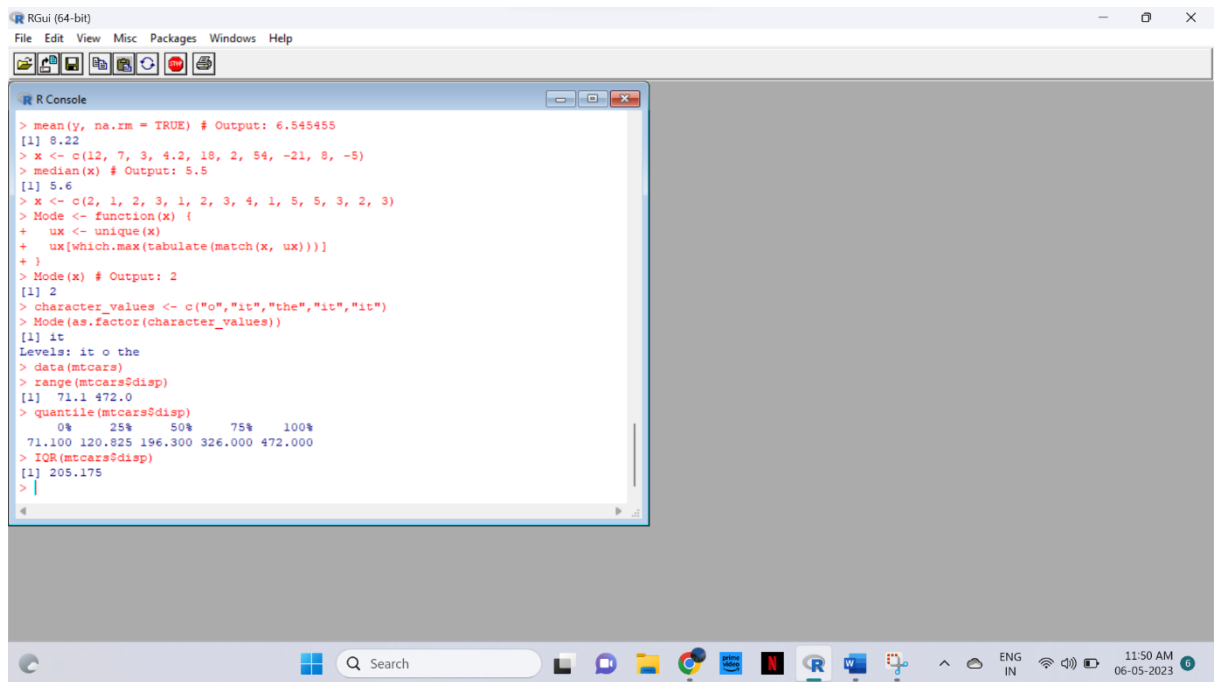
CODE

```
range(mtcars$disp)
```

```
range(mtcars$disp)
```

```
quantile(mtcars$disp)
```

```
IQR(mtcars$disp)
```



Exercise 7

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

CODE

library(dplyr)

data(mpg)

skew_city_mpg <- skewness(mpg\$cty, na.rm = TRUE)

skew_city_mpg

library(ggplot2)

qplot(mpg\$cty, geom = "histogram", binwidth = 2) +

labs(x = "City miles per gallon", y = "Frequency")

```
kurt_city_mpg <- kurtosis(mpg$cty, na.rm = TRUE)
```

```
kurt_city_mpg
```

