**Statistics with r programming LAB MANUAL\_-day 3**

**Aswini L**

**192124096**

**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**

**program:**

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

**mean(values)**

**output:**

**> # create a vector of values**

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

**>**

**> # calculate the mean of the values**

**> mean(values)**

**[1] 8.22**

**>**

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

**end.**

**PROGRAM:**

**# create a vector of values**

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

**# calculate the mean after trimming 3 values from each end**

**mean(values, trim = 0.3)**

**OUTPUT:**

**> # create a vector of values**

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

**>**

**> # calculate the mean after trimming 3 values from each end**

**> 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)**

**PROGRAM:**

**# create a vector of values**

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

**# calculate the mean of the values, excluding NA values**

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

**OUTPUT:**

**> # create a vector of values**

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

**>**

**> # calculate the mean of the values, excluding NA values**

**> 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:**

**# create a vector of values**

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

**# calculate the median of the values**

**median(values)**

**OUTPUT:**

**> # create a vector of values**

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

**>**

**> # calculate the median of the values**

**> 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) , (&quot;o&quot;,&quot;it&quot;,&quot;the&quot;,&quot;it&quot;,&quot;it&quot;)**

**PROGRAM:**

**# numeric data set**

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

**# user-defined function to calculate mode for numeric data set**

**mode\_numeric <- function(x) {**

**ux <- unique(x)**

**ux[which.max(tabulate(match(x, ux)))]**

**}**

**# calculate the mode of the numeric data set using the user-defined function**

**mode\_numeric(values\_numeric)**

**# character data set**

**values\_char <- c("o", "it", "the", "it", "it")**

**# calculate the mode of the character data set**

**mode(values\_char)**

**OUTPUT:**

**> # numeric data set**

**> values\_numeric <- c(2, 1, 2, 3, 1, 2, 3, 4, 1, 5, 5, 3, 2, 3)**

**>**

**> # user-defined function to calculate mode for numeric data set**

**> mode\_numeric <- function(x) {**

**+ ux <- unique(x)**

**+ ux[which.max(tabulate(match(x, ux)))]**

**+ }**

**>**

**> # calculate the mode of the numeric data set using the user-defined function**

**> mode\_numeric(values\_numeric)**

**[1] 2**

**>**

**>**

**> # character data set**

**> values\_char <- c("o", "it", "the", "it", "it")**

**>**

**> # calculate the mode of the character data set**

**> mode(values\_char)**

**[1] "character"**

**>**

**UNIVARIATE ANALYSIS IN R - MEASURES OF DISPERSION**

**Exercise: 4**

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

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

**PROGRAM:**

**# install and load the needed packages**

**install.packages("ggplot2") # for accessing the mpg dataset**

**library(ggplot2)**

**# download the mpg dataset from the URL and load it into R**

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

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

**max\_city\_mpg\_car <- mpg[mpg$cty == max(mpg$cty), "model"]**

**print(paste0("Car with maximum city miles per gallon: ", max\_city\_mpg\_car))**

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

**compact\_subcompact <- mpg[mpg$class %in% c("compact", "subcompact"),]**

**min\_disp\_cars <- compact\_subcompact[compact\_subcompact$displ == min(compact\_subcompact$displ), "model"]**

**print(paste0("Car(s) with minimum displacement in compact and subcompact class: ", min\_disp\_cars))**

**OUTPUT:**

**> # download the mpg dataset from the URL and load it into R**

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

**>**

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

**> max\_city\_mpg\_car <- mpg[mpg$cty == max(mpg$cty), "model"]**

**> print(paste0("Car with maximum city miles per gallon: ", max\_city\_mpg\_car))**

**[1] "Car with maximum city miles per gallon: new beetle"**

**>**

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

**> compact\_subcompact <- mpg[mpg$class %in% c("compact", "subcompact"),]**

**> min\_disp\_cars <- compact\_subcompact[compact\_subcompact$displ == min(compact\_subcompact$displ), "model"]**

**> print(paste0("Car(s) with minimum displacement in compact and subcompact class: ", min\_disp\_cars))**

**[1] "Car(s) with minimum displacement in compact and subcompact class: civic"**

**[2] "Car(s) with minimum displacement in compact and subcompact class: civic"**

**[3] "Car(s) with minimum displacement in compact and subcompact class: civic"**

**[4] "Car(s) with minimum displacement in compact and subcompact class: civic"**

**[5] "Car(s) with minimum displacement in compact and subcompact class: civic"**

**>**

**Exercise: 5**

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

**i) Find the standard deviation of city milles per gallon**

**ii) Find the variance of highway milles per gallon**

**PROGRAM:**

**# load the ggplot2 package**

**library(ggplot2)**

**# load the mpg dataset**

**data(mpg)**

**# i) Find the standard deviation of city miles per gallon**

**sd\_city\_mpg <- sd(mpg$cty)**

**print(paste0("Standard deviation of city miles per gallon: ", sd\_city\_mpg))**

**# ii) Find the variance of highway miles per gallon**

**var\_highway\_mpg <- var(mpg$hwy)**

**print(paste0("Variance of highway miles per gallon: ", var\_highway\_mpg))**

**OUTPUT:**

**# load the ggplot2 package**

**> library(ggplot2)**

**>**

**> # load the mpg dataset**

**> data(mpg)**

**>**

**> # i) Find the standard deviation of city miles per gallon**

**> sd\_city\_mpg <- sd(mpg$cty)**

**> print(paste0("Standard deviation of city miles per gallon: ", sd\_city\_mpg))**

**[1] "Standard deviation of city miles per gallon: 4.2559456788894"**

**>**

**> # ii) Find the variance of highway miles per gallon**

**> var\_highway\_mpg <- var(mpg$hwy)**

**> print(paste0("Variance of highway miles per gallon: ", var\_highway\_mpg))**

**[1] "Variance of highway miles per gallon: 35.4577785114266"**

**>**

**>**

**Exercise 6**

**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**

**PROGRAM:**

**# load the ggplot2 package**

**library(ggplot2)**

**# load the mpg dataset**

**data(mpg)**

**# i) Find the range of the disp in the data set mpg**

**range\_disp <- range(mpg$displ)**

**print(paste0("Range of disp in the mpg dataset: ", range\_disp))**

**# ii) Find the Quartile of the disp in the data set mpg**

**quantiles\_disp <- quantile(mpg$displ)**

**print(paste0("Quartiles of disp in the mpg dataset: ", quantiles\_disp))**

**# iii) Find the IQR of the disp column in the data set mpg**

**iqr\_disp <- IQR(mpg$displ)**

**print(paste0("IQR of disp in the mpg dataset: ", iqr\_disp))**

**OUTPUT:**

**> # load the ggplot2 package**

**> library(ggplot2)**

**>**

**> # load the mpg dataset**

**> data(mpg)**

**>**

**> # i) Find the range of the disp in the data set mpg**

**> range\_disp <- range(mpg$displ)**

**> print(paste0("Range of disp in the mpg dataset: ", range\_disp))**

**[1] "Range of disp in the mpg dataset: 1.6" "Range of disp in the mpg dataset: 7"**

**>**

**> # ii) Find the Quartile of the disp in the data set mpg**

**> quantiles\_disp <- quantile(mpg$displ)**

**> print(paste0("Quartiles of disp in the mpg dataset: ", quantiles\_disp))**

**[1] "Quartiles of disp in the mpg dataset: 1.6" "Quartiles of disp in the mpg dataset: 2.4"**

**[3] "Quartiles of disp in the mpg dataset: 3.3" "Quartiles of disp in the mpg dataset: 4.6"**

**[5] "Quartiles of disp in the mpg dataset: 7"**

**>**

**> # iii) Find the IQR of the disp column in the data set mpg**

**> iqr\_disp <- IQR(mpg$displ)**

**> print(paste0("IQR of disp in the mpg dataset: ", iqr\_disp))**

**[1] "IQR of disp in the mpg dataset: 2.2"**

**>**

**>**

**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**

**PROGRAM:**

**# load the ggplot2 package**

**library(ggplot2)**

**# load the mpg dataset**

**data(mpg)**

**# a. Find the skewness of city miles per mileage in the data set mpg**

**skewness\_city\_mpg <- skewness(mpg$cty)**

**print(paste0("Skewness of city miles per gallon in the mpg dataset: ", skewness\_city\_mpg))**

**# plot a histogram of the city miles per gallon column**

**qplot(mpg$cty, geom = "histogram", bins = 10,**

**main = "City Miles Per Gallon",**

**xlab = "Miles Per Gallon")**

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

**kurtosis\_city\_mpg <- kurtosis(mpg$cty)**

**print(paste0("Kurtosis of city miles per gallon in the mpg dataset: ", kurtosis\_city\_mpg))**

**OUTPUT:**

**[1] "Skewness of city miles per gallon in the mpg dataset: 1.75344843724469"**

**[1] "Kurtosis of city miles per gallon in the mpg dataset: 3.10207663098217**