**LAB-EXCERCISES**

**DT:10-02-23 ITA0443-STATISTICS WITH R PROGRAMMING**

**NAME:S.SHASHIKANTH**

**REG.NO:192125026**

**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**

**CODE:**

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

> x<-mean(a)

> print(x)

**OUTPUT:**

[1] 8.22

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

**end.**

**CODE:**

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

> sorted\_data <- sort(data)

> trimmed\_data <- sorted\_data[4:(length(sorted\_data)-3)]

> mean(trimmed\_data)

**OUTPUT:**

[1] 12.8

**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**

**CODE:**

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

> mean(data, na.rm = TRUE)

**OUTPUT:**

[1] 14.563636

**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**

**CODE:**

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

> median(data)

**OUTPUT:**

[1] 8

**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;)**

**CODE:**

> numeric\_data <- c(2,1,2,3,1,2,3,4,1,5,5,3,2,3)

> numeric\_mode <- names(which.max(table(numeric\_data)))

> numeric\_mode

**OUTPUT**:

[1] "2"

**CODE**:

> character\_data <- c("o","it","the","it","it")

> character\_mode <- names(which.max(table(character\_data)))

> character\_mode

**OUTPUT**:

[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**](https://vincentarelbundock.github.io/Rdatasets/datasets.html)

**Answer the following queries**

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

**CODE**:

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

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

**CODE:**

> 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**

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

**CODE:**

> 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

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

**CODE:**

> var(mpg$cty)

**OUTPUT:**

[1] 18.11307

**Exercise 3**

**Use the same dataset and perform the following queries**

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

CODE:

> range(mpg$displ)

OUTPUT:

[1] 1.6 7.0

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

CODE:

> quantile(mpg$displ)

OUTPUT:

0% 25% 50% 75% 100%

1.6 2.4 3.3 4.6 7.0

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

**CODE:**

> IQR(mpg$displ)

**OUTPUT**:

[1] 2.2

**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**

**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**

**CODE:**

library(ggplot2)

data("mpg")

skewness <- skew(mpg$cty)

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

qplot(mpg$cty, geom = "density", main = "Density Plot of City Miles per Gallon")

kurtosis <- kurtosis(mpg$cty)

cat("Kurtosis of city miles per gallon:", kurtosis, "\n")

**OUTPUT:**

Skewness of city miles per gallon: 0.3824208

Kurtosis of city miles per gallon: -0.9947507

**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**

**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 )**

**CODE:**

> dataframe1 <- data.frame(Reference = 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", "Male", "Male", "Female", "Female", "Female", "Male", "Female", "Female", "Female"),

+ TestNewOrFollowUp = c("Test1 New", "Test1 New", "Test2 New", "Test3 New", "Test1 New", "Test4 Follow-up", "Test4 New", "Test2 New", "Test3 New", "Test1 New"))

> xtabs(~Status + Gender + TestNewOrFollowUp, data = dataframe1)

**OUTPUT:**

Gender TestNewOrFollowUp Status

Female Test1 New Accepted 2

Rejected 1

Test3 New Accepted 1

Test4 Follow-up Accepted 1

Male Test1 New Accepted 1

Rejected 1

Test2 New Rejected 1

Test4 New Rejected 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.**

**CODE:**

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:**



**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.**

**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”.**

**CODE:**

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:**



**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**

**CODE::**

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

**OUTPUT:**



**5. Create a Boxplot graph for the relation between &quot;mpg&quot;(miles per galloon) and**

**&quot;cyl&quot;(number of Cylinders) for the dataset &quot;mtcars&quot; available in R Environment.**

**CODE:**

data(mtcars)

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

>

**OUTPUT:**

