**DAY-3 LAB 192125097**

**S.POOJITHA**

***1. 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***

**Ans:R Program**

a. # Create a vector.

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

# Find Mean.

result.mean <- mean(x)

print(result.mean)

b.# Create a vector.

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

# Find Trim

result.mean <- mean(x,trim = 0.3)

print(result.mean)

c.# Create a vector.

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

# Find mean.

result.mean <- mean(x)

print(result.mean)

# Find mean dropping NA values.

result.mean <- mean(x,na.rm = TRUE)

print(result.mean)

***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***

**Ans:R Program**

# Create the vector.

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

# Find the median.

median.result <- median(x)

print(median.result)

***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")***

**Ans:R Program**

# Create the function.

getmode <- function(v) {

uniqv <- unique(v)

uniqv[which.max(tabulate(match(v, uniqv)))]

}

# Create the vector with numbers.

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

# Calculate the mode using the user function.

result <- getmode(v)

print(result)

# Create the vector with characters.

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

# Calculate the mode using the user function.

result <- getmode(charv)

print(result)

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

***Answer the following queries***

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

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

**Ans:R Program**

mpg <- read.csv("C:/Users/Admin/Documents/R programming/day 3/mpg.csv")

# find car with maximum city MPG

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

mpg[max\_city\_mpg, "model"]

# find cars with minimum disp in compact and subcompact classes

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

min\_disp\_compact <- which.min(compact$displ[compact$class == "compact"])

min\_disp\_subcompact <- which.min(compact$displ[compact$class == "subcompact"])

# extract the corresponding car models

compact[min\_disp\_compact, "model"]

compact[min\_disp\_subcompact, "model"]

***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***

**Ans:R Program**

mpg <- read.csv

sd\_city <- sd(mpg$cty)

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

var\_highway <- var(mpg$hwy)

cat("Variance of highway miles per gallon: ", var\_highway, "\n")

***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***

**Ans:R Program**

mpg <- read.csv

disp\_range <- range(mpg$disp)

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

disp\_quartiles <- quantile(mpg$disp)

cat("Quartiles of disp: ", disp\_quartiles, "\n")

disp\_iqr <- IQR(mpg$disp)

cat("Interquartile range of disp: ", disp\_iqr, "\n")

***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***

***Use qplot function and display the graph for the city miles per mileage column***

**Ans:R Program**

data(mpg)

skewness(mpg$city)

qplot(mpg$city, geom = "histogram", binwidth = 1, xlab = "city miles per gallon")

kurtosis(mpg$city)

qplot(mpg$city, geom = "histogram", binwidth = 1, xlab = "city miles per gallon")

**Exercise: 8**

**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**

**R PROGRAM :**

data <- matrix(c(

"KRXH", "Accepted", "Female", "Test1", "New",

"KRPT", "Accepted", "Male", "Test1", "New",

"FHRA", "Rejected", "Male", "Test2", "New",

"CZKK", "Accepted", "Female", "Test3", "New",

"CQTN", "Rejected", "Female", "Test1", "New",

"PZXW", "Accepted", "Female", "Test4", "Follow-up",

"SZRZ", "Rejected", "Male", "Test4", "New",

"RMZE", "Rejected", "Female", "Test2", "New",

"STNX", "Accepted", "Female", "Test3", "New",

"TMDW", "Accepted", "Female", "Test1", "New"),

ncol = 5, byrow = TRUE)

dataframe1 <- data.frame(

Reference = data[, 1],

Status = data[, 2],

Gender = data[, 3],

TestNewOrFollowUp = data[, 5]

)

print(dataframe1)

status\_gender\_table <- table(dataframe1$Status, dataframe1$Gender)

print(status\_gender\_table)

**Exercise: 9**

**i) Use Two Categorical Variables and Discover the relationships within a**

**dataset**

**ii) Next, using the xtabs() function, apply two variables from “dataframe1 “, to**

**create a table delineating the relationship between the “Reference”**

**category, and the “Status” category.**

**iii) Save the file in the name of dataframe2**

**R PROGRAM :**

# Load the ggplot2 library

library(ggplot2)

# Plot a stacked bar chart of Status by Reference

ggplot(dataframe1, aes(x = Reference, fill = Status)) +

geom\_bar() +

labs(x = "Reference", y = "Count", title = "Status by Reference") +

theme\_bw()

# Create an example dataframe

Reference <- c("A", "A", "B", "B", "B", "C", "C", "C", "C")

Status <- c("Success", "Failure", "Failure", "Success", "Success", "Failure", "Success", "Success", "Success")

dataframe1 <- data.frame(Reference, Status)

# Create a contingency table using xtabs()

table1 <- xtabs(~ Reference + Status, data = dataframe1)

# Display the contingency table

table1

# Save the contingency table as a new file

dataframe2 <- as.data.frame(table1)

write.csv(dataframe2, "dataframe2.csv", row.names = FALSE)

**Exercise: 10**

**Use the same data frame using three Categorical Variables create a Multi-Dimensional Table**

**Apply three variables from “dataframe1” to create a Multi-Dimensional Cross-Tabulation of**

**“Status“, “Gender“, and “Test“.**

**R PROGRAM :**

# Create an example dataframe

Gender <- c("Male", "Male", "Female", "Female", "Male", "Female", "Male", "Female")

Test <- c("Test1", "Test2", "Test2", "Test3", "Test1", "Test2", "Test1", "Test3")

Status <- c("Success", "Failure", "Failure", "Success", "Success", "Failure", "Success", "Success")

dataframe1 <- data.frame(Gender, Test, Status)

# Create a multi-dimensional table using table() function

table1 <- table(dataframe1$Status, dataframe1$Gender, dataframe1$Test)

# Display the multi-dimensional table

table1

**Exercise: 11**

**Row Percentages**

**The R package “tigerstats” is required for the next two exercises.**

**1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.**

**2) Enclose the xtabs() formula in the tigerstats function, “rowPerc()” to display row**

**percentages for “Status” by “Test“.**

**R PRORAM :**

library(stringi)

library(tigerstats)

cross\_tab <- xtabs(~ Status + Reference, data = dataframe1)

row\_percentages <- rowPerc(cross\_tab)

print(row\_percentages)

**Exercise 12**

**Column Percentages**

**1) Create an xtabs() formula that cross-tabulates “Status“, and “Test“.**

**2) Enclose the xtabs() formula in the tigerstats function, “colPerc()” to display row**

**percentages for “Status” by “Test“.**

**R PROGRAM :**

cross\_tab <- xtabs(~ Status + Gender, data = dataframe1)

col\_percentages <- colPerc(cross\_tab)

print(col\_percentages)

**13. 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.**

**R PROGRAM :**

# Define the input vector and corresponding labels

values <- c(21, 62, 10, 53)

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

# Create the pie chart

pie(values, labels = labels, main = "City Pie-Chart")

# Add a legend at the top right corner of the chart

legend("topright", inset = 0.05, legend = labels, cex = 0.8, fill = rainbow(length(values)))

**14. 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.**

**R PROGRAM :**

# Install and load the required package

install.packages("plotrix")

library(plotrix)

# Create the dataset "political Knowledge"

political\_knowledge <- c("Low", "Medium", "High")

knowledge\_counts <- c(20, 30, 50)

# Set the colors for the pie chart

colors <- c("#FF5F5F", "#FFCC5F", "#5F9FFF")

# Create the 3D pie chart

pie3D(knowledge\_counts, labels = political\_knowledge, explode = 0.1,

main = "Political Knowledge", col = colors)

# Add a legend at the top right corner

legend("topright", legend = political\_knowledge, fill = colors, title = "Knowledge Level")

# Adjust the perspective of the chart

title(main = "Political Knowledge")

title(xlab = "")

title(ylab = "")

**15. 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”.**

**R PROGRAM :**

# create the vectors

H <- c(7, 12, 28, 3, 41)

M <- c("mar", "apr", "may", "jun", "jul")

# create the bar chart

barplot(H, names.arg = M, main = "Revenue chart")

**16. 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**

**R RPOGRAM :**

# Load the AirPassengers dataset

data(AirPassengers)

# Create a histogram with custom breaks

hist(AirPassengers, xlim = c(100, 700), breaks = seq(100, 700, by = 200),

main = "Histogram of AirPassengers", xlab = "Passengers", ylab = "Frequency")

**17. 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.**

**R PROGRAM :**

# load the mtcars dataset

data(mtcars)

# create a boxplot graph for mpg and cyl

boxplot(mpg ~ cyl, data = mtcars, xlab = "Number of Cylinders", ylab = "Miles per Gallon")