**ITA0448-R PROGRAMMING**

**NARESH GUPTHA**

**192124152**

**Day 3 assignment**

**1. Consider the data set occupationalStatus in the datasets package.**

**(a) What is the probability of a son having the same occupational status as his father? [Hint:**

**investigate what diag(x) does if x is a matrix.]**

**(b) Renormalize the data so that each row sums to 1. In the new data set the ith row**

**represents the conditional distribution of a son’s occupational status given that his father has**

**occupational status i.**

**(c) What is the probability that a son has occupational status between 1 and 3, given that his**

**father has status 1?**

**What if the father has occupational status 8?**

CODE:

a)data(occupationalStatus)

trans\_mat <- table(occupationalStatus$fath, occupationalStatus$son)

prob\_same <- sum(diag(trans\_mat))/sum(trans\_mat)

prob\_same

Output

[1] 0.2862006

b)trans\_mat\_norm <- trans\_mat / rowSums(trans\_mat)

prob\_son\_1\_3\_given\_fath\_1 <- sum(trans\_mat\_norm[1, 1:3])

prob\_son\_1\_3\_given\_fath\_1

output

[1] 0.5454545

c)prob\_son\_1\_3\_given\_fath\_8 <- sum(trans\_mat\_norm[8, 1:3])

prob\_son\_1\_3\_given\_fath\_8

output

[1] 0

**2. Create the following data frame, subsequently invert Gender for all individuals.**

**a) Name Age Height Weight Gender**

**Alex 25 177 57 M**

**Lilly 31 163 69 M**

**Mark 23 190 83 F**

**b) Create the below data frame**

**Name Working**

**Alex Yes**

**Lilly No**

**Mark No**

**c) Add the data frame column-wise to the previous one.**

**How many rows and columns does the new data frame have?**

CODE:

a)df <- data.frame(Name = c("Alex", "Lilly", "Mark"),

Age = c(25, 31, 23),

Height = c(177, 163, 190),

Weight = c(57, 69, 83),

Gender = c("M", "M", "F"))

df$Gender <- ifelse(df$Gender == "M", "F", "M")

b)working <- data.frame(Name = c("Alex", "Lilly", "Mark"),

Working = c("Yes", "No", "No"))

c)new\_df <- cbind(df, working$Working)

**3. A student recorded his/her scores on weekly R programming quizzes that were marked out**

**of a possible 10 points. His/Herscores were as follows:**

**8, 5, 8, 5, 7, 6, 7, 7, 5, 7, 5, 5, 6, 6, 9, 8, 9, 7, 9, 9, 6, 8, 6, 6, 7**

**What is the mode of his/her scores on the weekly R programming quizzes?**

CODE:

We can see that the score of 7 appears most frequently, with a total of 4 times. Therefore, the mode of the scores is 7.

**4. Construct the following data frame.**

**Countries population\_in\_million gdp\_per\_capita**

**A 100 2000**

**B 200 7000**

**C 120 15000**

**a) Write appropriate R code and reshape the above data frame from wide data format**

**to long data format.**

**b) Write R code and reshape from long to wide data format.**

CODE:

a)library(tidyr)

# Create data frame

df <- data.frame(Countries = c("A", "B", "C"),

population\_in\_million = c(100, 200, 120),

gdp\_per\_capita = c(2000, 7000, 15000))

# Reshape to long format

df\_long <- pivot\_longer(df, cols = c("population\_in\_million", "gdp\_per\_capita"),

names\_to = "Variable", values\_to = "Value")

b)# Reshape to wide format

df\_wide <- pivot\_wider(df\_long, names\_from = Variable, values\_from = Value)

**5. Consider the following data present. Create this file using windows notepad . Save the file**

**as input.csv using the save As All files(\*.\*) option in notepad.**

**i. Use appropriate R commands to read input.csv file.**

**ii. Analyze the CSV File and compute the following.**

**a. Get the maximum salary**

**b. Get the details of the person with max salary**

**c. Get all the people working in IT department**

**d. Get the persons in IT department whose salary is greater than 600**

**e. Get the people who joined on or after 2014**

**iii. Get the people who joined on or after 2014 and write the output onto a file called**

output.csv

CODE:

i)data <- read.csv("input.csv")

print(data)

Output:

id, name, salary, start\_date, dept

1 1 Rick 623.30 2012-01-01 IT

2 2 Dan 515.20 2013-09-23 Operations

3 3 Michelle 611.00 2014-11-15 IT

4 4 Ryan 729.00 2014-05-11 HR

5 NA Gary 843.25 2015-03-27 Finance

6 6 Nina 578.00 2013-05-21 IT

7 7 Simon 632.80 2013-07-30 Operations

8 8 Guru 722.50 2014-06-17 Finance

ii)data <- read.csv("input.csv")

print(is.data.frame(data))

print(ncol(data))

print(nrow(data))

Output:

[1] TRUE

[1] 5

[1] 8

a)# Create a data frame.

data <- read.csv("input.csv")

# Get the max salary from data frame.

sal <- max(data$salary)

print(sal)

Output:

[1] 843.25

b)# Create a data frame.

data <- read.csv("input.csv")

# Get the max salary from data frame.

sal <- max(data$salary)

# Get the person detail having max salary.

retval <- subset(data, salary == max(salary))

print(retval)

Output:

id name salary start\_date dept

5 NA Gary 843.25 2015-03-27 Finance

c)# Create a data frame.

data <- read.csv("input.csv")

retval <- subset( data, dept == "IT")

print(retval)

Output:

id name salary start\_date dept

1 1 Rick 623.3 2012-01-01 IT

3 3 Michelle 611.0 2014-11-15 IT

6 6 Nina 578.0 2013-05-21 IT

d)# Create a data frame.

data <- read.csv("input.csv")

info <- subset(data, salary > 600 & dept == "IT")

print(info)

Output:

id name salary start\_date dept

1 1 Rick 623.3 2012-01-01 IT

3 3 Michelle 611.0 2014-11-15 IT

e)# Create a data frame.

data <- read.csv("input.csv")

retval <- subset(data, as.Date(start\_date) > as.Date("2014-01-01"))

print(retval)

Output:

id name salary start\_date dept

3 3 Michelle 611.00 2014-11-15 IT

4 4 Ryan 729.00 2014-05-11 HR

5 NA Gary 843.25 2015-03-27 Finance

8 8 Guru 722.50 2014-06-17 Finance

iii)

Code:

# Create a data frame.

data <- read.csv("input.csv")

retval <- subset(data, as.Date(start\_date) > as.Date("2014-01-01"))

# Write filtered data into a new file.

write.csv(retval,"output.csv")

newdata <- read.csv("output.csv")

print(newdata)

Output:

X id name salary start\_date dept

1 3 3 Michelle 611.00 2014-11-15 IT

2 4 4 Ryan 729.00 2014-05-11 HR

3 5 NA Gary 843.25 2015-03-27 Finance

4 8 8 Guru 722.50 2014-06-17 Finance