**1. Use the package RcmdrPlugin.IPSUR. data(RcmdrTestDrive) and perform the below operations:**

**a. Calculate the average salary by gender and smoking status.**

***Ans:***

data(RcmdrTestDrive)

library(RcmdrPlugin.IPSUR)

View(RcmdrPlugin.IPSUR)

library(dplyr)

RcmdrTestDrive %>%

group\_by(Smoking\_Status = RcmdrTestDrive$smoking, Gender = RcmdrTestDrive$gender) %>%

summarise(mean\_salary = mean(salary))

**b. Which gender has the highest mean salary?**

***Ans:***

library(RcmdrPlugin.IPSUR)

View(RcmdrPlugin.IPSUR)

library(dplyr)

Highest\_Mean\_Salary <- as.data.frame(RcmdrTestDrive %>%

group\_by(Gender = RcmdrTestDrive$gender) %>%

summarise(mean\_salary = mean(salary)))

Highest\_Mean\_Salary[which.max(Highest\_Mean\_Salary$mean\_salary),]

# Gender Male has the highest mean salary

**c. Report the highest mean salary.**

***Ans:***

library(RcmdrPlugin.IPSUR)

View(RcmdrPlugin.IPSUR)

library(dplyr)

RcmdrTestDrive %>%

group\_by(Smoking\_Status = RcmdrTestDrive$smoking, Gender = RcmdrTestDrive$gender) %>%

summarise(mean\_salary = mean(salary)) %>% arrange(desc(mean\_salary))

# The highest mean salary(which is of Gender = Male as per above answer) is 743.3.

**d. Compare the spreads for the genders by calculating the standard deviation of salary by gender.**

***Ans:***

library(RcmdrPlugin.IPSUR)

library(dplyr)

RcmdrTestDrive %>%group\_by(Gender = RcmdrTestDrive$gender) %>%

summarise(sd\_salary = sd(salary))

# Table for the standard deviation of salary on the basis of Gender.

Gender sd\_salary

1 Female 131.

2 Male 159

# The Standard Deviation of the Male mean salary is more than the Female mean salary. Hence the spread of the Male salary

is more than the Female salary.