**Programming Codes:**

#Reading Insurance Data

Insdata <- read.csv("SwedishMotorInsurance.csv")

head(Insdata)

#1. The committee is interested to know each field of the data collected through descriptive analysis to gain basic insights into the data set and to prepare for further analysis.

summary(Insdata)

#2. The total value of payment by an insurance company is an important factor to be monitored. So, the committee has decided to find whether this payment is related to number of claims and the number of insured policy years. They also want to visualize the results for better understanding.

#Approach 1 - liner regression model

lm1<-lm(Insdata$Payment~Insdata$Claims+Insdata$Insured)

summary(lm1)

#Approach 2 - Find correlation between variables

cor(Insdata$Payment,Insdata$Claims)

plot(Insdata$Payment,Insdata$Claims, col = "red",xlab="Claims",ylab="Payment",cex=0.6,

main="Claims Vs.Payment",cex.main=0.8,cex.axis=0.6 )

cor(Insdata$Payment,Insdata$Insured)

plot(Insdata$Payment,Insdata$Insured, col = "green", xlab="Insured Amount",ylab="Payment",cex=0.6,

main="Insured Amount Vs.Payment",cex.main=0.8,cex.axis=0.6)

#3. The committee wants to figure out the reasons for insurance payment increase and decrease. So, they have decided to find whether distance, location, bonus, make, and insured amount or claims are affecting the payment or all or some of these are affecting it.

lm2<-lm(Insdata$Payment~., data = Insdata)

summary(lm2)

#4. The insurance company is planning to establish a new branch office, so they are interested to find at what location, kilo-meter, and bonus level their insured amount, claims, and payment get increased. (Hint: Aggregate Dataset)

ZoneResult <- apply(Insdata[,c(5,6,7)],2, function(x)tapply(x, Insdata$Zone, mean))

ZoneResult

KmResult <- apply(Insdata[,c(5,6,7)],2, function(x)tapply(x, Insdata$Kilometres, mean))

KmResult

BonusResult <- apply(Insdata[,c(5,6,7)],2, function(x)tapply(x, Insdata$Bonus, mean))

BonusResult

#5. The committee wants to understand what affects their claim rates so as to decide the right premiums for a certain set of situations. Hence, they need to find whether the insured amount, zone, kilo meter, bonus, or make affects the claim rates and to what extent.

md <- lm(Insdata$Claims ~ Insdata$Kilometres + Insdata$Zone + Insdata$Bonus + Insdata$Make + Insdata$Insured)

summary(md)

-------------------------------------------------------------------The End-------------------------------------------------------------------