**Analyze the report of Swedish Motor Insurance**

**CODES :**

Ins<-read.csv(file.choose(),header=T)

**Q1**

head(Ins)

#to know each field of data collected through descriptive analysis

#to gain basics insight into data set

summary(Ins)

**Q2**

MDLlm<-lm(Ins$Payment~Ins$Claims+Ins$Insured)

MDLlm

summary(MDLlm)

#as per the pvalue is less,payment is related to no.of claim & no of insured policy per year

#visualisation

plot(Ins$Payment,Ins$Claims,col="red")

plot(Ins$Payment,Ins$Insured,col="blue")

**Q3**

MDL1lm2<-lm(Ins$Payment~.,data=Ins)

MDL1lm2

summary(MDL1lm2)

##results

##increase or decrease of payment related to

#distance,insured,claims,less related to location

**Q4**

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

grpzone

#results

#zone 4 has highest nos of claims, payments & insured policy

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

grpkm

#result

#claims and payments have highest value at kilometer group 2

#the insured no of years has highest value at group 1

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

grpbns

#result

#bonus group 7 has maximum insured policies, claims & payments.

**Q5**

MDL5lm5<-lm(Claims~Kilometres+Zone+Bonus+Make+Insured,data=Ins)

summary(MDL5lm5)

#result

#zone,bonus,make,insured policies affected highly to claims, distance affected less.

**Solution**

**1.**

The results provide the minimum and maximum values. It also provides the mean and median values of all variables. From this you can understand the spread of data. We can see that claims and payment also have null or zero values, however the insured column does not have a zero value. This specifies that there are few entries where the car has been insured for a given period of time. However, no claim or payment has been made for that combination of car make, zone, and kilometres.

**2**.

As per the value of payment by an insurance company is an important factor to be monitored, so it found that payment is related to no. of claims & no of insured policies years. Visualisation occurred for better understanding.

**3.**

The reason for insurance payments increase & decrease depends on distance, insured policies per year, claims and less depends on location.

**4.**

A –zone 1 has more insured years.

B-km group 2 has max payments though are insured lesser no of tears than grp 1 & even have higher claims than grp1.

C-zone 4 has highest claims hence higher payments.

D-bonus grp 7 has high amounts of insured years, payments, claims.

**5.**

P values are high, hence assume that km, zone, insured, bonus, make are making impact on claims.