#Swedish Motor insurance Project

#Load Libraries

library(MASS)

#Read Data

InsData <- read.csv("C:/analytics/Insurance/SwedishMotorInsurance.csv")

#1 Data Analysis

summary(InsData)

#2 Correlation analysis of Payment ,Insured,Claims

pairs(~Payment+Insured+Claims ,data=InsData )

#Relation of payment with Categorical Variables

prop.table(table( InsData$Kilometres ) )

prop.table(table( InsData$Zone ) )

prop.table(table( InsData$Bonus ))

prop.table(table( InsData$Make ))

#3 Model for Payment

Rel <- lm(InsData$Payment~InsData$Kilometres+InsData$Zone+InsData$Bonus+InsData$Make+InsData$Insured+InsData$Claims )

summary(Rel)

#4 Group by Distance/ZOne/Bonus

#By Dist

InsSub <- subset(InsData , Select=C(

FDist<- factor(InsData$Kilometres)

InsDist <- subset(InsData, select= c (Insured,Claims,Payment,Kilometres) )

AggrbyKM <- aggregate(.~FDist , InsDist ,sum)

AggrbyKM$Claimratio <- AggrbyKM$Claims/AggrbyKM$Insured \*100

AggrbyKM$CostPerInsured <-AggrbyKM$Payment /AggrbyKM$Insured

head(AggrbyKM)

summary(AggrbyKM)

#by Zone

FZone<- factor(InsData$Zone)

InsZone <- subset(InsData, select= c (Insured,Claims,Payment,Zone) )

AggrbyZone <- aggregate(.~FZone , InsZone ,sum)

AggrbyZone$Claimratio <- AggrbyZone$Claims/AggrbyZone$Insured \*100

AggrbyZone$CostPerInsured <-AggrbyZone$Payment /AggrbyZone$Insured

head(AggrbyZone)

summary(AggrbyZone)

#By bonus

FBonus<- factor(InsData$Bonus)

InsBonus <- subset(InsData, select= c (Insured,Claims,Payment,Bonus) )

AggrbyBonus <- aggregate(.~FBonus , InsBonus ,sum)

AggrbyBonus$Claimratio <- AggrbyBonus$Claims/AggrbyBonus$Insured \*100

AggrbyBonus$CostPerInsured <-AggrbyBonus$Payment /AggrbyBonus$Insured

head(AggrbyBonus)

summary(AggrbyBonus)

#5 Model for Claims

str(InsData)

ModelClaims <- lm(InsData$Claims~InsData$Kilometres+InsData$Zone+InsData$Bonus+InsData$Make+InsData$Insured)

summary(ModelClaims)