##Read data file

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

View(Hospital)

hist(Hospital$AGE,

main="Frequency of Patients",

col ="green",

xlab = "Age")

attach(Hospital)

AGE <-as.factor(AGE)

summary(AGE)

View(summary(AGE))

##Aggregate function is used to add the expenditure from each age

aggregate(TOTCHG~AGE,FUN = sum,data = Hospital)

##Max

max(aggregate(TOTCHG~AGE,FUN = sum,data = Hospital))

hist(APRDRG,

col = "green",

main = "Frequency of Treatments",

xlab = "Treatment Categories")

APRDRG\_fact <-as.factor(Hospital$APRDRG)

summary(APRDRG\_fact)

#Max

which.max(summary(APRDRG\_fact))

df<-aggregate(TOTCHG~APRDRG,FUN = sum,data = Hospital)

df

#Total charge

df[which.max(df$TOTCHG),]

#First remove "NA" value

Hospital<-na.omit(Hospital)

#Factorize the Race variable

Hospital$RACE<-as.factor(Hospital$RACE)

#ANOVA function with TOTCHG and RACE Variable.

model\_aov<-aov(TOTCHG~RACE,data = Hospital)

#ANOVA RESULTS

model\_aov

summary(model\_aov)

#Getting max hospital cost per race

summary(Hospital$RACE)

#Analyze the severity of costs

Hospital$FEMALE<-as.factor(Hospital$FEMALE)

#calling Regression function

model\_lm4 <-lm(TOTCHG~AGE+FEMALE,data = Hospital)

summary(model\_lm4)

#comparing genders

summary(Hospital$FEMALE)

#Linier Regression

Hospital$RACE<-as.factor(Hospital$RACE)

model\_lm5<-lm(LOS~AGE+FEMALE+RACE,data = Hospital)

summary(model\_lm5)

#Linier Regressin

model\_lm6<-lm(TOTCHG~AGE+FEMALE+RACE+LOS+APRDRG,data = Hospital)

summary(model\_lm6)