```
#PROJECT 7 - HEALTHCARE COST ANALYSIS#

library(openxlsx)
hospitalcosts_R<-read.xlsx(file.choose())
View(hospitalcosts_R)
head(hospitalcosts_R)

#To record the patient statistics, the agency wants to find the age category of people
#who frequently visits the hospital and has the maximum expenditure.

hist(hospitalcosts_R$AGE,main = "frequency of patient",col="green",xlab ="Age")
summary(as.factor(hospitalcosts_R$AGE))

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

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

#In order of severity of the diagnosis and treatments and to find out the expensive treatments,
#the agency wants to find the diagnosis related group that has maximum hospitalization and expenditure.

hist(hospitalcosts_R$APRDRG,main = "frequency of treatments",col="blue",xlab="diagnosis related group ")

APRDRG_fact<-as.factor(hospitalcosts_R$APRDRG)
summary(APRDRG_fact)
which.max(summary(APRDRG_fact))

df<-aggregate(TOTCHG~APRDRG,FUN = sum,data=hospitalcosts_R)
df
df[which.max(df$TOTCHG),]

#To make sure that there is no malpractice, the agency needs to analyze if the race
#of the patient is related to the hospitalization costs.

hospitalcosts_R<-na.omit(hospitalcosts_R)
hospitalcosts_R$RACE<-as.factor(hospitalcosts_R$RACE)
summary(RACE)

model_aov<-aov(TOTCHG~RACE,data = hospitalcosts_R)
model_aov
summary(model_aov)

#To properly utilize the costs, the agency has to analyze the severity of the hospital
#costs by age and gender for proper allocation of resources.

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

model_lm4<-lm(TOTCHG~AGE+FEMALE,data = hospitalcosts_R) #calling Regression funtion

summary(model_lm4)

summary(hospitalcosts_R$FEMALE)

#Since the length of stay is the crucial factor for inpatients,
#the agency wants to find if the length of stay can be predicted from age, gender, and race.
```

```
model_lm5<-lm(LOS~AGE+FEMALE+RACE,data= hospitalcosts_R)
summary(model_lm5)

#To perform a complete analysis, the agency wants to find the variable that mainly affects the hospital costs.

model_lm6<-lm(TOTCHG~AGE+FEMALE+RACE+LOS+APRDRG,data=hospitalcosts_R)
summary(model_lm6)
```