```
setwd('C:/Users/Admin/Desktop')
getwd()
df <- read excel("Healthcare Data.xlsx")</pre>
View(df)
str(df)
summary(df)
# To record the patient statistics, the agency wants to find the age category of
people who frequent the hospital and has the maximum expenditure
head(df)
summary(df)
head(df$AGE)
summary(df$AGE)
table(df$AGE)
hist(df$AGE)
summary(as.factor(df$AGE))
max(table(df$AGE))
max(summary(as.factor(df$AGE)))
which.max(table(df$AGE))
age <- aggregate (TOTCHG ~ AGE, data = df, sum)
max(age)
#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.
t <- table(df$APRDRG)
d <- as.data.frame(t)</pre>
names(d)[1] = 'Diagnosis Group'
which.max(table(df$APRDRG))
which.max(t)
```

```
race of the patient is related to the hospitalization costs
table(df$RACE)
df$RACE <- as.factor(df$RACE)</pre>
fit <- lm(TOTCHG ~ RACE, data=df)</pre>
fit
summary(fit)
fit1 <- aov(TOTCHG ~ RACE, data=df)</pre>
summary(fit1)
df <- na.omit(df)</pre>
#To properly utilize the costs, the agency has to analyze the severity of the
hospital costs by age and gender for proper allocation of resources.
table(df$FEMALE)
a <- aov(TOTCHG ~ AGE+FEMALE, data=df)</pre>
summary(a)
b <- lm(TOTCHG ~ AGE+FEMALE, data=df)</pre>
summary(b)
#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.
table(df$LOS)
cat <- aov(LOS ~ AGE+FEMALE+RACE, data=df)</pre>
summary(cat)
cat <- lm(LOS ~ AGE+FEMALE+RACE, data=df)</pre>
summary(cat)
#To perform a complete analysis, the agency wants to find the variable that
mainly affects the hospital costs.
aov(TOTCHG ~., data=df)
mod <- lm(TOTCHG ~ .,data=df)</pre>
summary (mod)
```

#To make sure that there is no malpractice, the agency needs to analyze if the