Health care Project in R

Business Scenario:

A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyze the data to research on the healthcare costs and their utilization.

Attributes Description:

AGE - Age of the patient discharged

FEMALE - Binary variable that indicates if the patient is female

LOS - Length of stay, in days

RACE - Race of the patient (specified numerically)

TOTCHG - Hospital discharge costs

APRDRG - All Patient Refined Diagnosis Related Groups

Expected Outcome:

1) To record the patient statistics, the agency wants to find the age category of people who frequent the hospital and has the maximum expenditure.

2) 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.

3) 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.

4) 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.

5) 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.

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

R Codes:

hospital <- read.csv(file.choose())

View(hospital)

str(hospital)

length(hospital)

head(hospital)

summary(hospital)

attach(hospital)

table(AGE)

hist(AGE)

as.factor(AGE)

age\_max <- aggregate(TOTCHG~AGE,FUN=sum,data=hospital)

age\_max[which.max(age\_max$TOTCHG),]

table(hospital$APRDRG)

which.max(summary(as.factor(hospital$APRDRG)))

diagno\_cost <- aggregate(TOTCHG~APRDRG,FUN=sum,data=hospital)

diagno\_cost[which.max(diagno\_cost$TOTCHG),]

class(RACE)

summary(as.factor(RACE))

na.omit(hospital$RACE)

Model <- aov(TOTCHG ~ RACE, data = hospital)

summary(Model)

plot(Model)

model2 <- lm(TOTCHG ~ AGE + FEMALE, data = hospital)

summary((model2))

plot(model2)

Model3 <- lm(LOS~AGE+FEMALE+RACE)

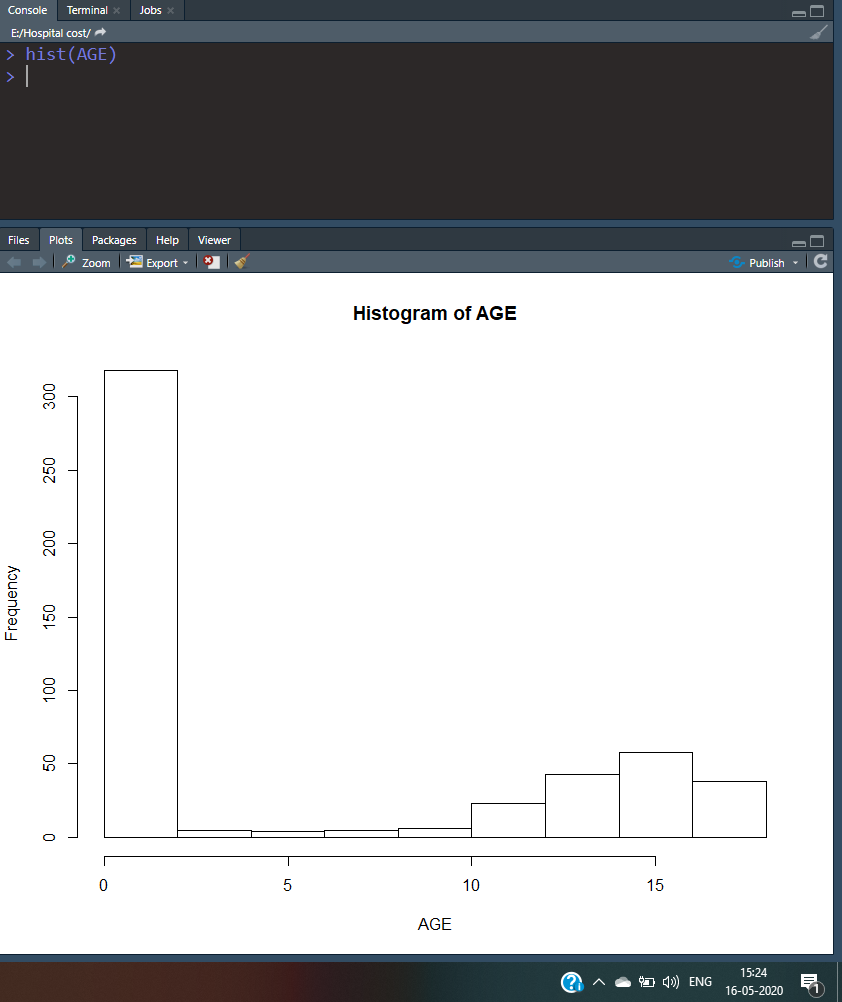
summary(Model3)

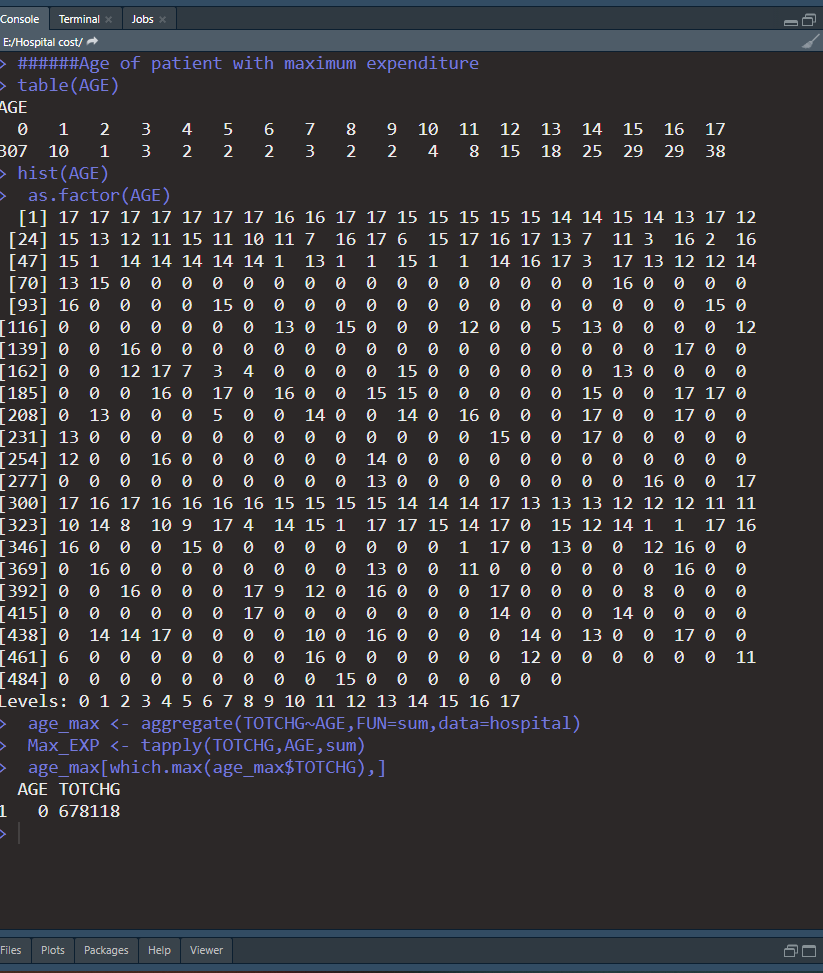
Final\_Model <- lm(TOTCHG~.,data=hospital)

summary(Final\_Model)

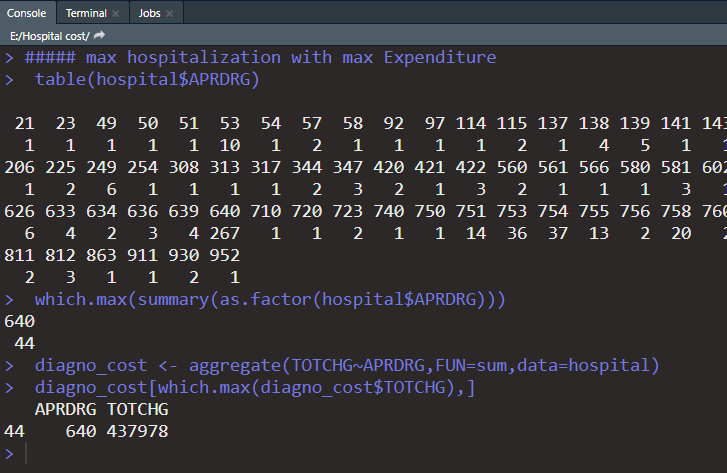
Output:

1)Age of patient with maximum expenditure-





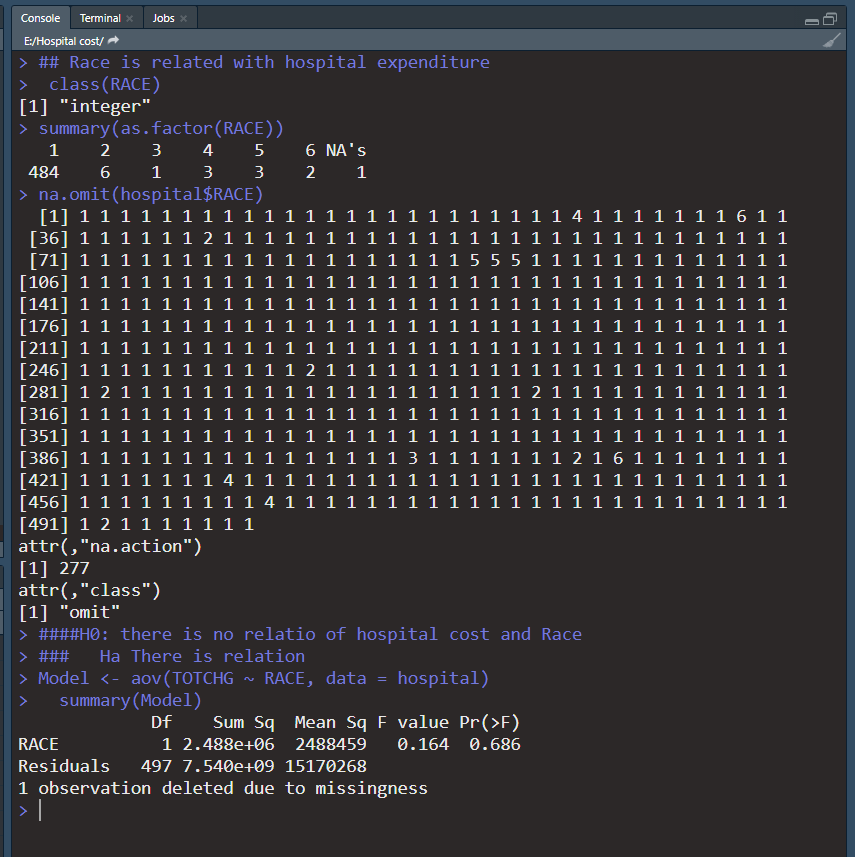
2) Max hospitalization with max Expenditure-

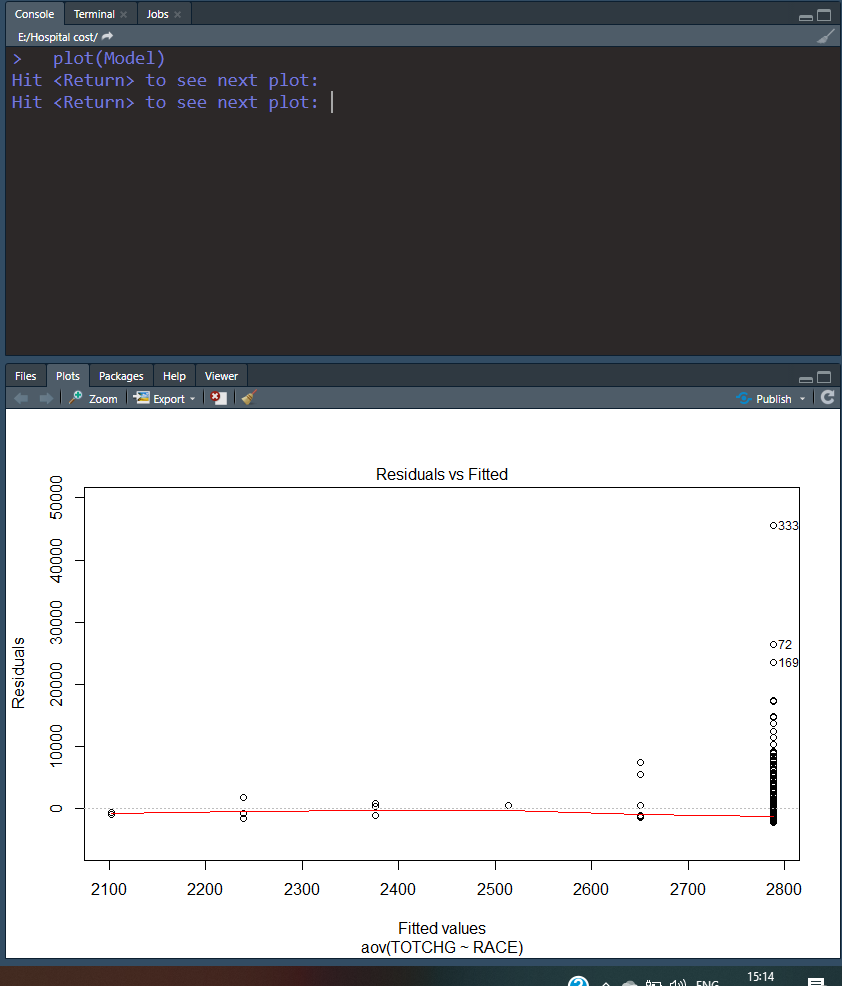


3) The Race is related with hospital expenditure-

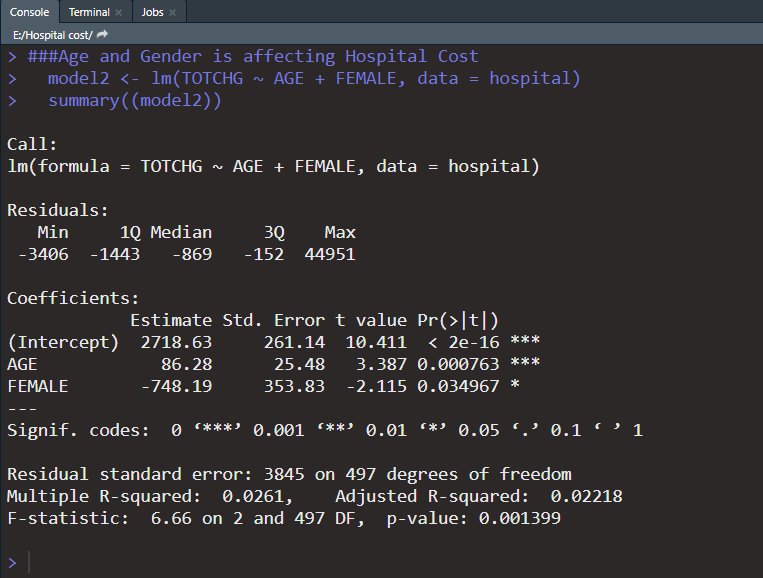
H0: there is no relation of hospital cost and Race

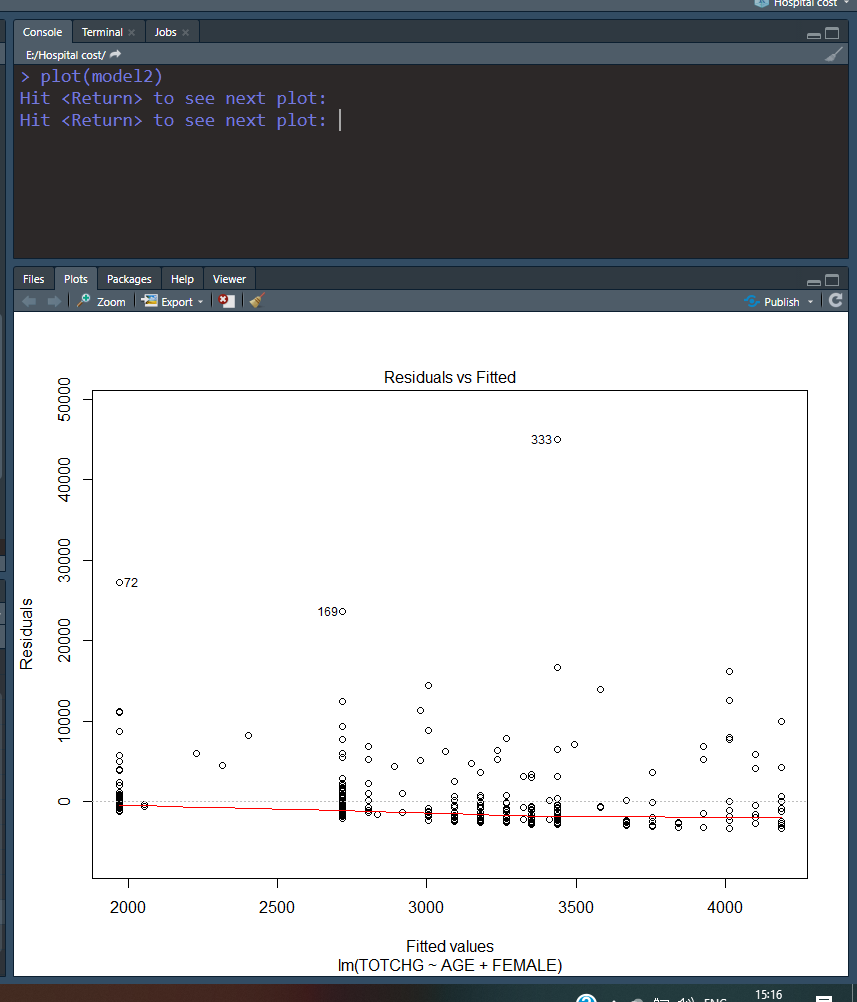
Ha: There is relation



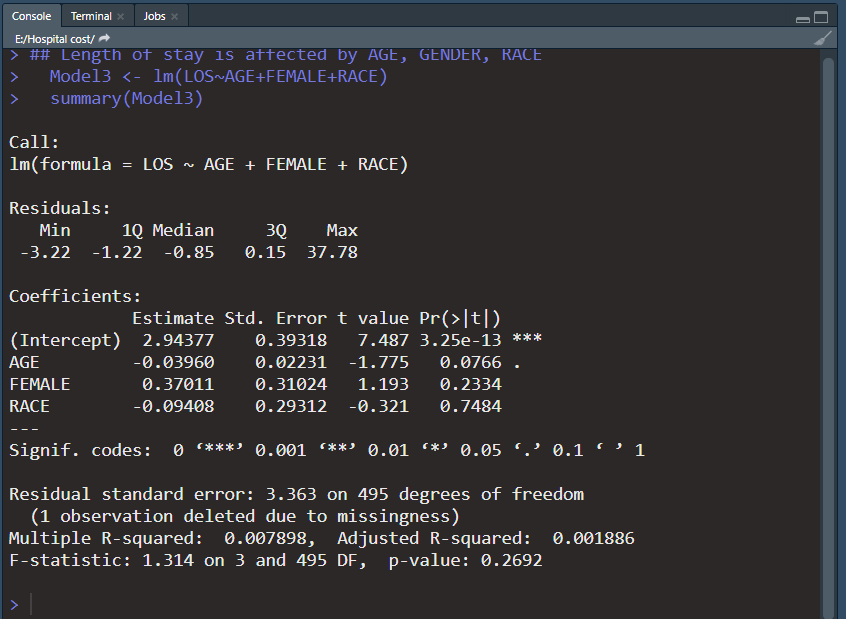


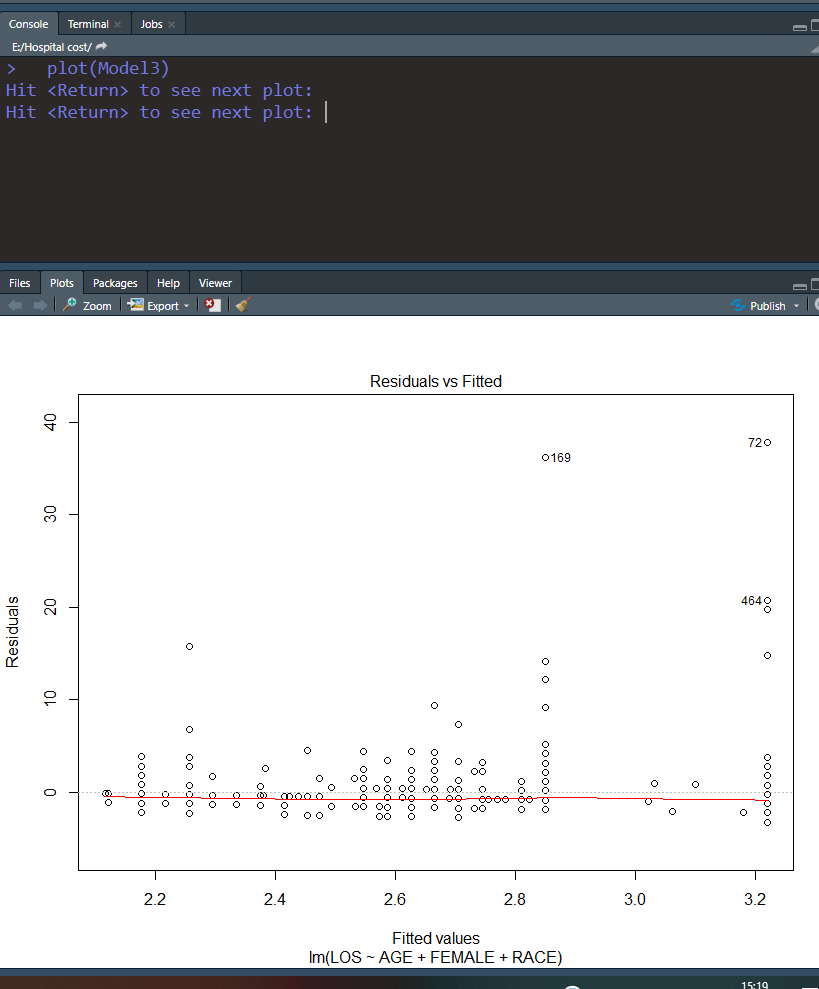
4) Age and Gender is affecting Hospital Cost-



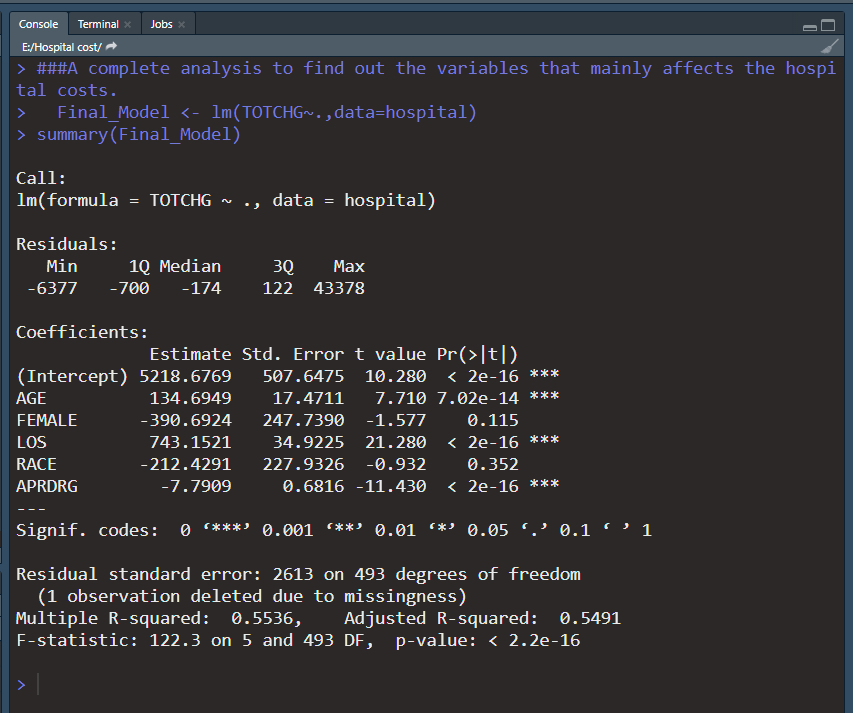


5) Length of stay is affected by AGE, GENDER, RACE-





6) A complete analysis to find out the variables that mainly affects the hospital costs-



Analysis:

1. 0 age group(infants) are mostly visiting hospital with max expenditure 678118.
2. 640 category of diagnosis Refined group has maximum expenditure of 437978
3. p value is higher than .05 so we accept the null hypothesis means there is no significant relation between Race and hospital cost.
4. Age and gender are significant in the model so it plays important role in determining hospital cost. Though R2 and R2 is not high so model is not fit to be good.
5. There is no relation between Length of Stay and AGE, FEMALE and RACE.
6. Age, Length of stay and diagnostic refined group affects the total Hospital cost.
7. Hospital Cost is having positive relation with Length of stay and Age and negative relation with diagnostic refined group affects.