**PROBLEM STATEMENT:** 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.

**R CODE:**

setwd("D:/Data Science with R")

hospital\_DS <- read.csv("HospitalCosts.csv")

View(hospital\_DS)

# Question No 1

DF = as.data.frame(table(hospital\_DS$AGE))

DF2 = aggregate(TOTCHG~AGE, hospital\_DS, mean)

Final\_DF = cbind(DF2$AGE, DF$Freq, DF2$TOTCHG)

colnames(Final\_DF) <- c("Age", "# of patients (PART A)", "Expenditure (PART B)")

View(Final\_DF)

# Question No 2

diagnosis\_cost = aggregate(TOTCHG~APRDRG+LOS, hospital\_DS, mean)

library(dplyr)

diagnosis\_cost = arrange(diagnosis\_cost,desc(TOTCHG))

colnames(diagnosis\_cost) <- c("APRDRG", "AVG LOS", "AVG Expenditure")

View(diagnosis\_cost)

# Question NO 3

AOV\_hosp = aov(TOTCHG~RACE,hospital\_DS)

summary(AOV\_hosp)

# Question No 4

hospital\_costs = aggregate(TOTCHG~AGE+FEMALE, hospital\_DS, mean)

colnames(hospital\_costs) <- c("AGE", "FEMALE FLAG", "AVG Expenditure")

View(hospital\_costs)

# Question No 5

hosp\_lm = lm(LOS~AGE+FEMALE+RACE,hospital\_DS)

hosp\_lm

summary(hosp\_lm)

# Question No 6

hosp\_lm2= lm(TOTCHG~., hospital\_DS)

hosp\_lm2

summary(hosp\_lm2)

# Drop FEMALE and RACE as they are not significant variables

hosp\_lm2 = lm(TOTCHG~AGE+LOS+APRDRG, hospital\_DS)

summary(hosp\_lm2)

# Make prediction

Prediction = predict(hosp\_lm2, hospital\_DS)

hosp\_Final = cbind(hospital\_DS,Prediction)

View(hosp\_Final)

hosp\_Error\_pct = transform(hosp\_Final, Error\_pct=abs(TOTCHG-Prediction)/TOTCHG)

View(hosp\_Error\_pct)

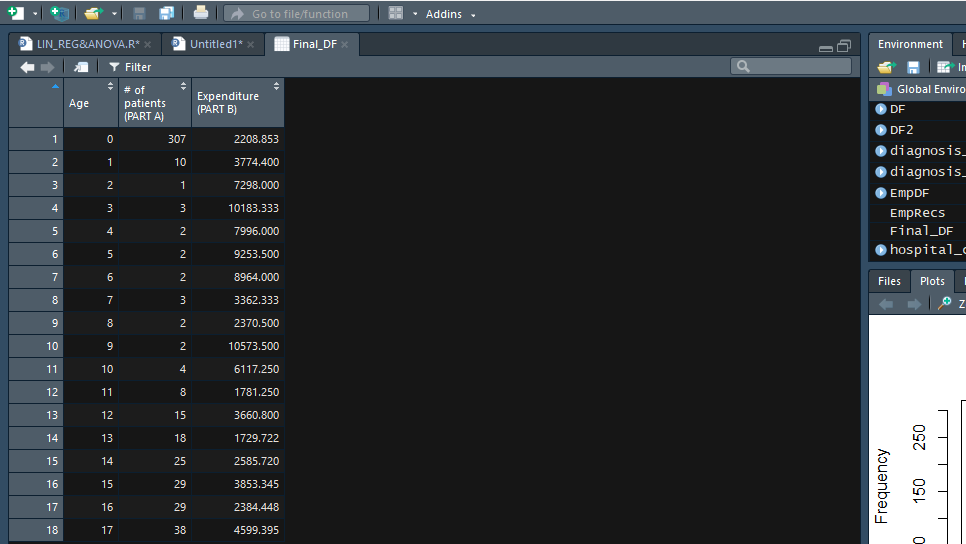
OverallErrorRate = mean(hosp\_Error\_pct$Error\_pct)

OverallErrorRate

AccuracyRate = 1- OverallErrorRate

AccuracyRate

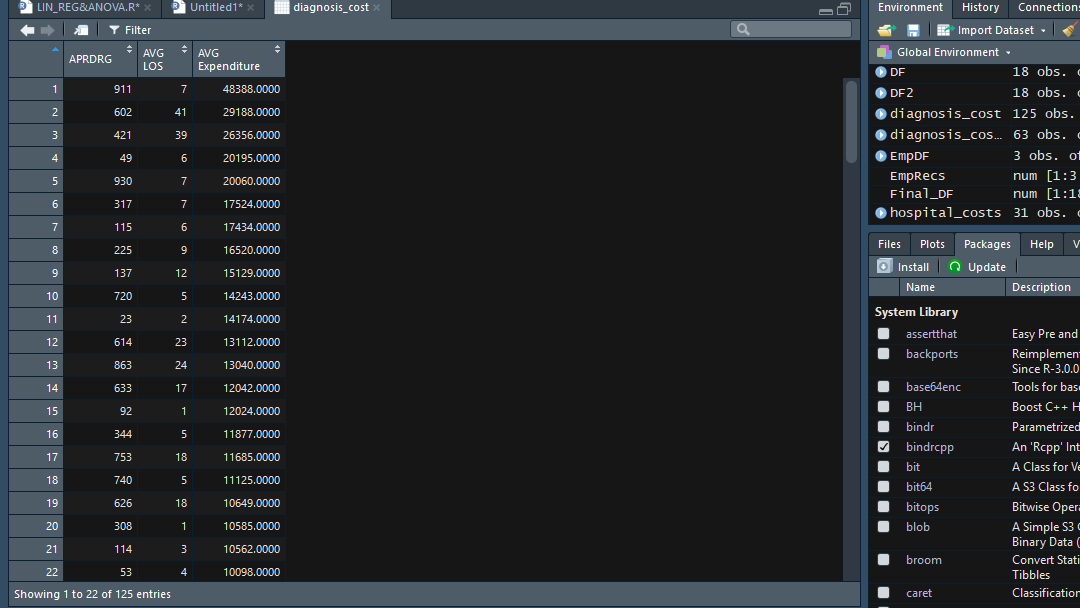
**QUESTION NO 1:**



**INSIGHT:**

**The age category 0 years (307) frequent the hospital and has the maximum expenditure (2208.8).**

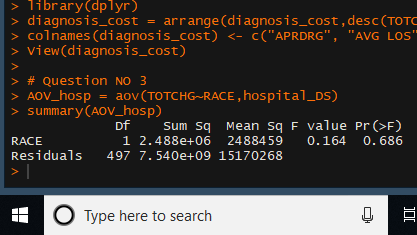
**QUESTION NO 2:**



**INSIGHT:**

**911 (APRDRG) with LOS 7 has the maximum average expenditure – 48388**

**QUESTION NO 3:**



**INSIGHT:**

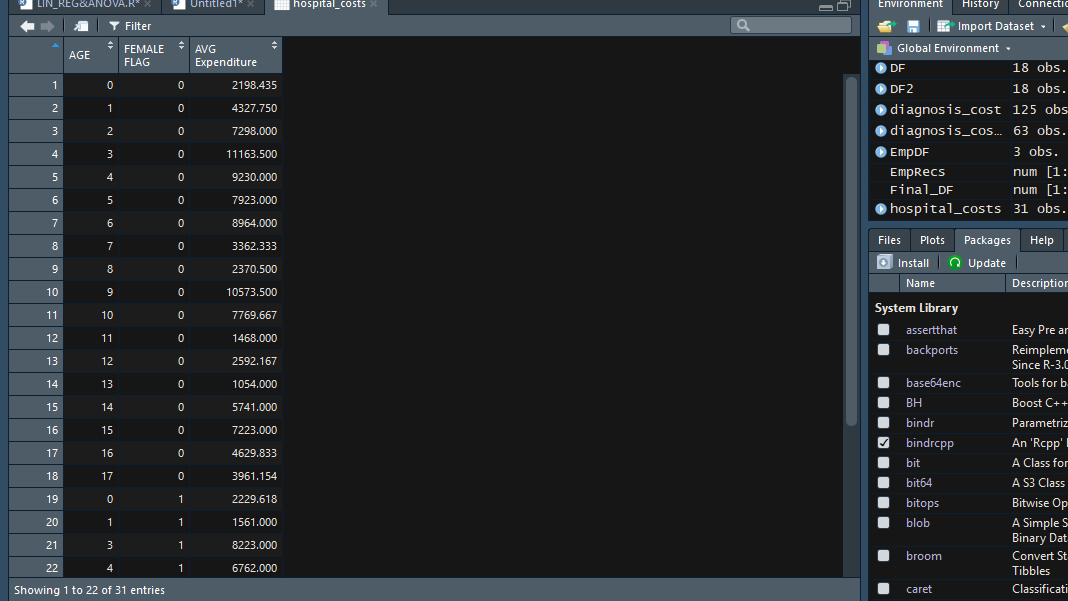
**H0 - mean RACE 1 = mean RACE 2 =mean RACE 3 = mean RACE 4 =mean RACE 5 =mean RACE 6**

**Ha – Either of them is different from each other**

**From the output probability p > 5%, H0 cannot be rejected**

**So there is no malpractice, the race of the patient is NOT related to the hospitalization costs**

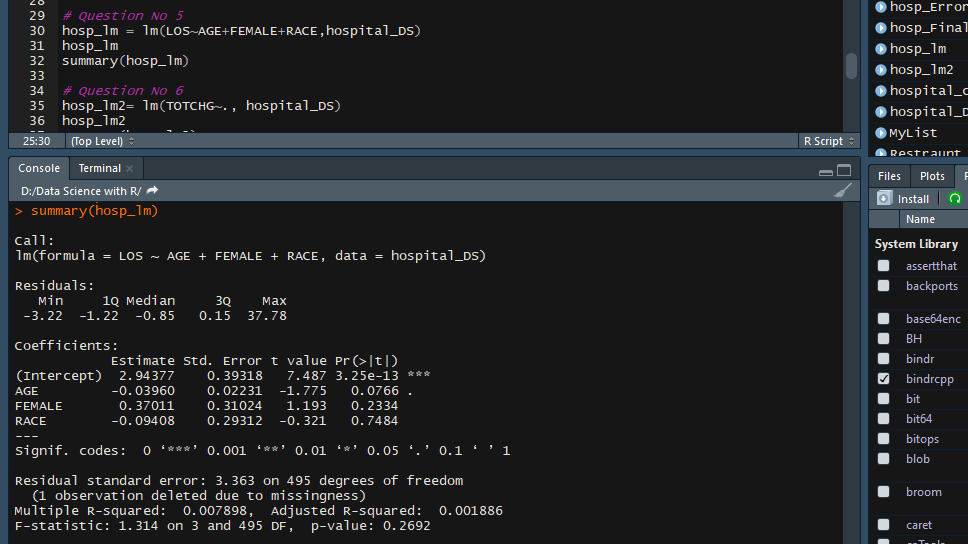
**QUESTION NO 4:**



**INSIGHT:**

**The severity of the hospital costs by age and gender is obtained which helps in proper allocation of resources.**

**QUESTION NO 5:**

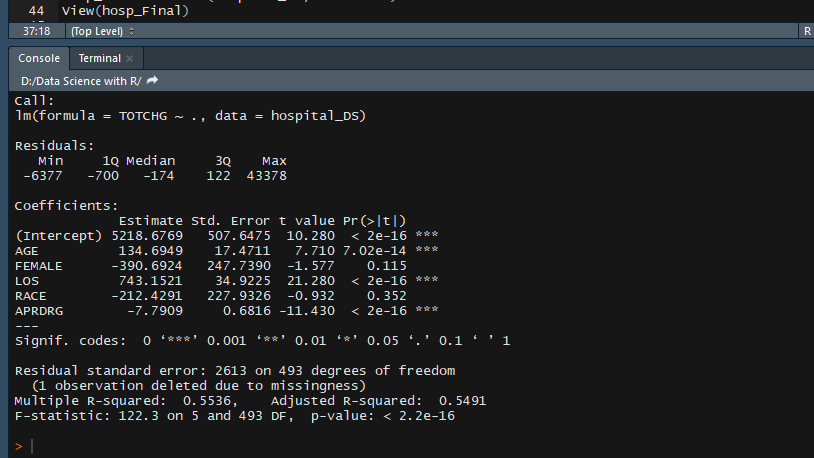
****

**INSIGHT:**

**From the output we can infer that no variables are significant.**

**So length of stay CANNOT be predicted from age, gender, and race.**

**QUESTION NO 6:**

****

**INSIGHT:**

**Based on the output we can say that AGE, LOS and APRDRG affects the total Hospital costs.**