Code for question 2

########AS Group Assignment##########

#The null hypothesis for ANOVA is that the mean

#(average value of the dependent variable) is the same for all groups.

# The alternative #or research hypothesis is that the average is not the

# same for all groups.

set.seed(123)

setwd("G:/Grreat lakes/Advance stat/Group Assignment")

mydata<- read.csv("Expectations Evaluation.csv")

mydata$Y2<- as.numeric(mydata$Y2)

mydata$Y1<- as.numeric(mydata$Y1)

#glm(Y1~.,data=mydata)

# Outlier Analysis - Varaiable

outlier\_upper=function(x){

q = quantile(x)

names(q) = NULL

q1 = q[2]

q3 = q[4]

QR = q3-q1

return(q3+1.5\*QR);

}

outlier\_lower=function(x){

q = quantile(x)

names(q) = NULL

q1 = q[2]

q3 = q[4]

QR = q3-q1

return(q1-1.5\*QR);

}

# outlier limits validation ------------------

Y2\_upper = outlier\_upper(mydata$Y2)

Y2\_lower = outlier\_lower(mydata$Y2)

Y1\_upper = outlier\_upper(mydata$Y1)

Y1\_lower = outlier\_lower(mydata$Y1)

# Outlier data

mydata[mydata$Y2>Y2\_upper | mydata$Y2<Y2\_lower , ]

mydata[mydata$Y1>Y1\_upper | mydata$Y1<Y1\_lower , ]

mydata = subset( mydata, mydata$Y2<=Y2\_upper & mydata$Y2>=Y2\_lower)

mydata = subset( mydata, mydata$Y1<=Y1\_upper & mydata$Y1>=Y1\_lower)

nrow(mydata)

#mydata$Expectatations.Manipulation=ifelse(mydata$Expectatations.Manipulation=="h",7,1)

#mean(HI[,1])

#mean(HI[,2])

aov(Y1~Quality.Manipulation,mydata)->ft

summary(ft)

TukeyHSD(ft)

tab<-table(mydata$Y1,mydata$Quality.Manipulation)

tab

mn\_b<-1+24+54+88+135+90+42+16

mn\_g<-6+24+72+105+120+161+16

mn\_b

mn\_g

#420,504

#mean(tab[,1])

#mean(tab[,2])

aov(Y1~Expectatations.Manipulation,mydata)->ft

summary(ft)

TukeyHSD(ft)

tab<-table(mydata$Y1,mydata$Expectatations.Manipulation)

tab

mn\_h<- 1+20+48+72+125+90+98+8

mn\_l<-10+30+88+115+120+105+24

mn\_h

mn\_l

aov(Y1~Expectatations.Manipulation+Quality.Manipulation,mydata)->ft

summary(ft)

aov(Y1~Expectatations.Manipulation\*Quality.Manipulation,mydata)->ft

summary(ft)

TukeyHSD(ft)

mydata$Expectatations.Manipulation=ifelse(mydata$Expectatations.Manipulation=="h",7,1)

ft1<- lm(Expectatations.Manipulation~Quality.Manipulation,data=mydata)

summary(ft1)

aov(Expectatations.Manipulation~Quality.Manipulation,mydata)->ft

summary(ft)

TukeyHSD(ft)

mydata$Expectatations.Manipulation=ifelse(mydata$Expectatations.Manipulation==7,"h",1)

mydata$random <- runif(nrow(mydata), 0, 1);

mydata <- mydata[order(mydata$random),]

mydata.dev <- mydata[which(mydata$random <= 0.75),]

mydata.val <- mydata[which(mydata$random > 0.75),]

c(nrow(mydata.dev), nrow(mydata.val))

ft1<- lm(Y1~Quality.Manipulation,data=mydata.dev)

#ft1<- lm(Y1~.,data=mydata.dev)

str(mydata.val$Quality.Manipulation)

summary(ft1)

Quality.Manipulation<-data.frame(Quality.Manipulation= mydata.val$Quality.Manipulation)

str(Quality.Manipulation)

predict(ft1 , c(Quality.Manipulation))

mydata.val$predicted\_score<-predict(ft1 , c(Quality.Manipulation))

View(mydata.val)

plot(ft1)

mydata$Expectatations.Manipulation=ifelse(mydata$Expectatations.Manipulation=="h",7,1)

ft1<- lm(Expectatations.Manipulation~Quality.Manipulation,data=mydata)

summary(ft1)

plot(ft1)

cor(mydata$Y1,mydata$Y2)

plot(mydata$Y1,mydata$Y2)

table(mydata$Y1,mydata$Y2)