# linear regression using diabetic data(BMI vs DBP or PGC)

View(Diabatic\_data1)

data1<-Diabatic\_data1

View(data1)

set.seed(2)

# sample.split function is present in this package

library(caTools)

#we divide the data with ratio 0.7

split<- sample.split(data1$HasRiskDiabetes, SplitRatio = 0.8)

# we will split data into training and testing data set

training\_data1<-subset(data1, split="TRUE") # as TRUE so training data set

testing\_data1<- subset(data1, split="FALSE") # as FALSE so testing data set

# see the summary of data1

summary(data1)

# data visualization

plot(data1$BodyMassIndex, data1$DiastolicBloodPressureInMMHG)

# Build model lm<- for linear regression (. means all variable will include)

data1\_model<- lm(data1$HasRiskDiabetes~., data = training\_data1) # DiastolicBloodPressureInMMHG -> dependent var

summary(data1\_model) # if p value not <0.05% then remove that variable

# predict the value

data1\_predict<- predict(data1\_model, newdata = testing\_data1)

data1\_predict

TAB<-table(testing\_data1$HasRiskDiabetes, data1\_predict>0.05)

TAB

# FALSE TRUE

#0 81 419

#1 3 265

(81+265)/(81+3+419+265)

# 0.4505208 = 45% acqurate

table(Actualvalue=testing\_data1$HasRiskDiabetes, PredictedvalueBymodel=data1\_predict> 0.05)

# acquracy of the model

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# To compare predicted values and actual(test\_value) values, we can use plot

#plot(testing\_data1$BodyMassIndex, type = "l", lty=1.8, col="green")

#lines(data1\_predict, type = "l", col="blue")

#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

plot(data1$BodyMassIndex, data1$DiastolicBloodPressureInMMHG)

# to draw regression line

plot(data1$BodyMassIndex, data1$DiastolicBloodPressureInMMHG)

abline(lm(data1$BodyMassIndex~data1$DiastolicBloodPressureInMMHG), col="red")

# to see corplot

cr<- cor(data1)

library(corrplot)

corrplot(cr, type = "lower")

corrplot(cr, method = "number")