Assignment 10

Durga Pokharel

18/01/2022

#[Q.N.1] Create two quantitative variables: age and body-mass index (BMI) with random samples of size 1000 each: \* Age: 0 to 99 (random samples) \* BMI: 10 to 40 (random samples)

Age

age <- sample(0:99,size = 1000, replace =T)

BMI

bmi <- sample(10:40, size = 1000, replace = T)

# [Q.N.2] Create a binary variable sex (1=Male and 0=Female) of 1000 random samples

# [Q.N.3] Create a data frame as df containing four variables/features:Serial Number, BMI, Age and Sex

sn <- c(1:1000)  
df <- data.frame(sn,bmi,age,sex)  
head(df)

## sn bmi age sex  
## 1 1 29 93 0  
## 2 2 12 79 1  
## 3 3 28 13 0  
## 4 4 10 4 1  
## 5 5 25 28 0  
## 6 6 22 97 1

# [Q.N.4] For replication of the results, use your class roll number as random.seed during analysis

# [Q.N.5] Split the data into “train” and “test” data using 80-20 partition

set.seed(11)  
nd = sample(2,nrow(df), replace= T,prob = c(0.8,0.2))  
train\_data <- df[nd ==1,]  
test\_data <- df[nd==2, ]

# [Q.N,6] Fit a linear regression model with BMI as dependent variable and age and sex and predictors in the train data samples

lm1 <- lm(bmi~.,data = train\_data)  
summary(lm1)

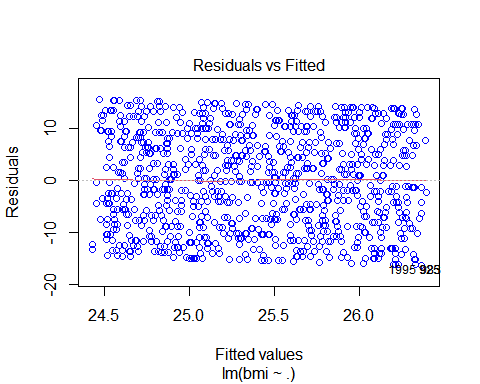
##   
## Call:  
## lm(formula = bmi ~ ., data = train\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -16.3629 -7.7817 -0.0603 8.0123 15.5321   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 25.4806795 0.6492593 39.246 <2e-16 \*\*\*  
## sn 0.0009306 0.0007817 1.190 0.2341   
## age -0.0012126 0.0080016 -0.152 0.8796   
## sex -0.9622368 0.4547669 -2.116 0.0345 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.081 on 1597 degrees of freedom  
## Multiple R-squared: 0.003692, Adjusted R-squared: 0.00182   
## F-statistic: 1.972 on 3 and 1597 DF, p-value: 0.1162

Here p value grater than 0.05. Hence dependent variable normailly distributed.

# [Q.N.7] Conduct residual analysis of the fitted model with graphs (suggestive) and tests (confirmative)

Linearity of resudial **Graphical (Suggestive)**

plot(lm1, which = 1, col= c("blue"))



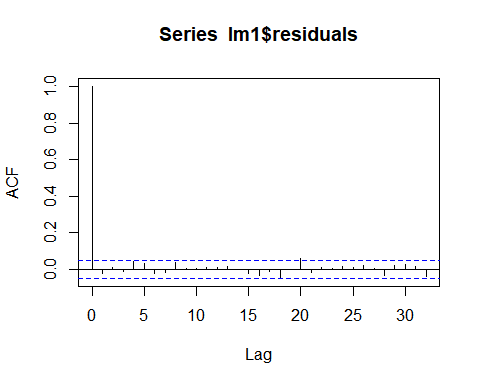
**Calculation (Confirmative)**

summary(lm1$residuals)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -16.36293 -7.78170 -0.06028 0.00000 8.01231 15.53215

Independence **Graphical (suggestive)**

acf(lm1$residuals)



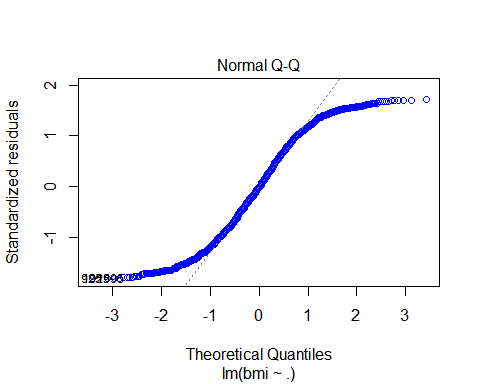
**Calculatin(Confirmative)**

library(car)  
durbinWatsonTest(lm1)

## lag Autocorrelation D-W Statistic p-value  
## 1 -0.02486152 2.049292 0.3  
## Alternative hypothesis: rho != 0

Normality of Residuals **Graphical(suggestive)**

plot(lm1, which = 2, col = c("blue"))

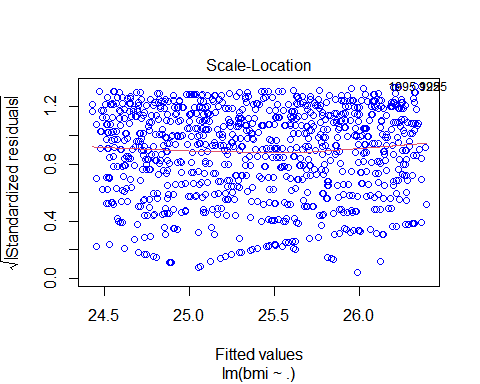
 **Calculation(Confirmative)**

shapiro.test(lm1$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: lm1$residuals  
## W = 0.95206, p-value < 2.2e-16

Equal Variance **Graphical(Suggestive)**

plot(lm1,which = 3, col= c("blue"))



**Calculation(Confirmative)**

library(lmtest)

## Warning: package 'lmtest' was built under R version 4.1.2

bptest(lm1)

##   
## studentized Breusch-Pagan test  
##   
## data: lm1  
## BP = 5.8142, df = 3, p-value = 0.121

# [Q.N.8] Use the fitted model to predict the random test data samples

pred <- predict(lm1, data = test\_data)  
length(pred)

## [1] 1601

length(test\_data$bmi)

## [1] 399

# [Q.N.9] Get R-square, MSE and RMSE for training as well as test data and interpret them carefully

library('caret')

## Warning: package 'caret' was built under R version 4.1.2

data.frame(R2 = R2(pred,   
train\_data$bmi),  
RMSE = RMSE(pred,   
train\_data$bmi),  
MAE = MAE(pred,   
train\_data$bmi))

## R2 RMSE MAE  
## 1 0.00369164 9.069891 7.866808

From this value of r-square only 0.45% data fit the regression model. Which is actually low. Similarly values for RMSE and MAE is respectively 8.909007 and 7.745684. It is better to have low error. Also, I have tried to get value of R2, RMSE and MAE for test data I got error Error in cor(obs, pred, use = ifelse(na.rm, "complete.obs", "everything")) : incompatible dimensions I tried to solve but I could not able.