R Code for
Linear Regression Analysis and its Diagnostics

```
# To fit a simple regression model y on x
x <- c(0,0.75,1.5,2.25,3,3.75,4.5,5.25,6,6.75,7.5,8.25,9,9.75,10.5,11.25,12,12.75)
y <- c(54.3,50.8,58,54.6,45.3,47,51.7,43.3,44.7,38.5,42.1,40,32,34.6,32.8,33.4,28.7,26.9)
z <- cbind(y,x)
z
plot(x,y)
cor(x,y)
cor.test(x,y)

hist(y)
boxplot(y)
M1<- lm(y~x)
ss <- summary(M1) # To print summary of results of fitted line
ss

#ANOVA Table
Aov <- aov (y~x)
summary(Aov)

# For regression diagnostics
resi<-residuals(Aov)  # To find residual values
y.hat<- fitted.values(Aov) # To find fitted values of y

# Checking for independence of observations i.e. autocorrelation
library(MASS)
str <- studres(Aov)
k<-1:18                    #As the length of x or y is 18
plot(k, str)

# Checking for Constant variance assumption
plot(y.hat,resi)
abline(h=0)

# Checking for normality of residuals
qqnorm(resi)
qqline(resi)
hist(resi)

# To merge the above 3 assumptions in one graph
library(e1071)
```

```
library(lattice)
library(PASWR)
checking.plots(Aov)

# Test for autocorrelation
library(lmtest)
dwtest(M1)  # Durban Watson test

# Breusch-Pagan Test
#H0: Homoscedasticity. The errors have constant variance about the true model.
#H1: Heteroscedasticity. The errors have non-constant variance about the true model.
#install.packages(lmtest)
library(lmtest)
bptest(M1)

# Test for normality of residulas
shapiro.test(resi)# If number of observations < 50  else
ks.test(resi,"pnorm", m= mean(resi), sd= sd(resi))

# To fit a Multiple Linear Regression Model
x <- c(0,0.75,1.5,2.25,3,3.75,4.5,5.25,6,6.75,7.5,8.25,9,9.75,10.5,11.25,12,12.75)
y <- c(54.3,50.8,58,54.6,45.3,47,51.7,43.3,44.7,38.5,42.1,40,32,34.6,32.8,33.4,28.7,26.9)
z <- c(18,16,9,11,9,7,4,3,8,5,1,4,3,4.6,8,4,7,9)
z1 <- cbind(z,x,y)
z1

hist(y)
boxplot(y)
M2<-lm(y~x+z)        #multiple linear regression
ss<-summary(M2)      # To print summary of results of fitted line
ss

# For regression diagnostics
resi<-residuals(M1)  # To find residual values
y.hat<- fitted.values(M1) # To find fitted values of y

Aov <- aov(y~x+z)
# Checking for independence of errors i.e. Autocorrelation
library(MASS)
sr<-studres(Aov)  #students's residual
k<-1:18
plot(k, sr)
```

```
# Checking for Constant variance assumption i.e. homoscedasticity
plot(y.hat,resi)  #if shape is of cone then variance is not constant.

# Checking for normality of errors
qqnorm(resi)
qqline(resi)
hist(resi)

# To merge the above 3 assumptions in one graph
library(e1071)
library(lattice)
library(PASWR)
checking.plots(Aov)

#test for checking auotcorrelation i.e. observations are independent
install.packages(lmtest)
library(lmtest)
dwtest(M1)  # durban watson test

# Breusch-Pagan Test
#H0: Homoscedasticity. The errors have constant variance about the true model.
#H1: Heteroscedasticity. The errors have non-constant variance about the true model.
#install.packages(lmtest)
library(lmtest)
bptest(M1)

# Test for normality of residulas
shapiro.test(resi)# If number of observations < 50  else
ks.test(resi,"pnorm", m= mean(resi), sd= sd(resi))

#test for checking multicollinearity i.e. VIF is
#install.packages("mctest")
library(mctest)
mctest(z1[,-1],z1[,1],type ="i")   #  only for independent variables
```