Library(dplyr)

Library(corpcor)

Library(mctest)

Library(car)

Library(readxl)

**Data name:**

Data(‘ ‘)

Data = dataframe( )

View( )

Read data:

Read.csv( )

Read\_excel( )

**Model Diagnostic Checking: Normality and Heteroskedesticity**

par(mfrow=c(2,2))

plot(regression)

**Pairwise Correlation among independent variables :**

library(GGally)

X<- excel\_file\_name[,2:11]

ggpairs(X)

AND

library(corpcor)

cor2pcor(cov(X))

here, X = independent variables

**Diagnostic Checking for Multi-Collinearity: Farrar Glauber Test**

Library(mctest)

Omcdiag(regression)

Imcdiag(regression)

* For this degrees of freedom at 5% level of significance, the theoretical value of F is 1.89774. Thus, the F test shows that either the variable ‘experience’ or ‘age’ or ‘education’ will be the root cause of multicollinearity.
* Though the F -value for ‘education’ is also significant, it may happen due to inclusion of highly collinear variables such as ‘age’ and ‘experience’.

**Ppcor package to compute the partial correlation along with t-statistic and corresponding p-values:**

install.packages("ppcor")

library(ppcor)

pcor(X, method = "pearson")

X = independent variables

**Checking VIF values:**

Library(car)

Vif(regress)

**Auto correlation:**

x = seq(1,length(reg$residuals))

p = ggplot() + geom\_line(data = hseinv, aes(x = x, y = reg$residuals), color = "red") + geom\_hline(yintercept=0, linetype="dashed", color='blue')

print(p)

where reg is the regression.

One way to solve problems in time series data, is to include more independent variables.

The inclusion of these variables will help to remove auto-correlation between the variables.

Box plot

Ggplot(data = data\_name, aes( x = data\_name$x, y = data\_name$y)) + geom\_boxplot( )