Multicollinearity Test

Suryanshu

6 February 2018

# Loading of Data

library(MASS)  
#reading the dataset into data  
data <- Boston  
str(data)

## 'data.frame': 506 obs. of 14 variables:  
## $ crim : num 0.00632 0.02731 0.02729 0.03237 0.06905 ...  
## $ zn : num 18 0 0 0 0 0 12.5 12.5 12.5 12.5 ...  
## $ indus : num 2.31 7.07 7.07 2.18 2.18 2.18 7.87 7.87 7.87 7.87 ...  
## $ chas : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ nox : num 0.538 0.469 0.469 0.458 0.458 0.458 0.524 0.524 0.524 0.524 ...  
## $ rm : num 6.58 6.42 7.18 7 7.15 ...  
## $ age : num 65.2 78.9 61.1 45.8 54.2 58.7 66.6 96.1 100 85.9 ...  
## $ dis : num 4.09 4.97 4.97 6.06 6.06 ...  
## $ rad : int 1 2 2 3 3 3 5 5 5 5 ...  
## $ tax : num 296 242 242 222 222 222 311 311 311 311 ...  
## $ ptratio: num 15.3 17.8 17.8 18.7 18.7 18.7 15.2 15.2 15.2 15.2 ...  
## $ black : num 397 397 393 395 397 ...  
## $ lstat : num 4.98 9.14 4.03 2.94 5.33 ...  
## $ medv : num 24 21.6 34.7 33.4 36.2 28.7 22.9 27.1 16.5 18.9 ...

# Collinearity Test using VIF

library(usdm)

## Warning: package 'usdm' was built under R version 3.3.3

## Loading required package: sp

## Loading required package: raster

## Warning: package 'raster' was built under R version 3.3.3

##   
## Attaching package: 'raster'

## The following objects are masked from 'package:MASS':  
##   
## area, select

df = data  
vif(df)

## Variables VIF  
## 1 crim 1.831537  
## 2 zn 2.352186  
## 3 indus 3.992503  
## 4 chas 1.095223  
## 5 nox 4.586920  
## 6 rm 2.260374  
## 7 age 3.100843  
## 8 dis 4.396007  
## 9 rad 7.808198  
## 10 tax 9.205542  
## 11 ptratio 1.993016  
## 12 black 1.381463  
## 13 lstat 3.581585  
## 14 medv 3.855684

#If VIF > 4.0 then I generally assume multicollinearity remove all those   
#Predictor Variables before fitting them into my model

# collinearity Testr using omcdiag

library(mctest)

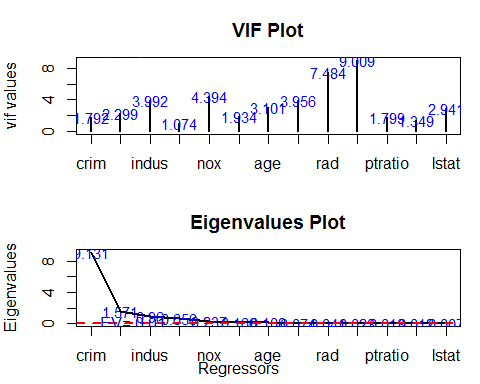
## Warning: package 'mctest' was built under R version 3.3.3

x <- data[,-14]  
y <- data[,14]  
  
omcdiag(x,y, detr=0.001, conf=0.99)

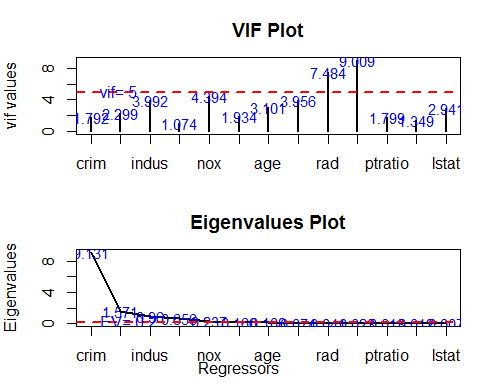
##   
## Call:  
## omcdiag(x = x, y = y, detr = 0.001, conf = 0.99)  
##   
##   
## Overall Multicollinearity Diagnostics  
##   
## MC Results detection  
## Determinant |X'X|: 0.0001 1  
## Farrar Chi-Square: 4486.4729 1  
## Red Indicator: 0.4432 0  
## Sum of Lambda Inverse: 45.1229 0  
## Theil's Method: -1.2643 0  
## Condition Number: 87.3183 1  
##   
## 1 --> COLLINEARITY is detected by the test   
## 0 --> COLLINEARITY is not detected by the test  
##   
## ===================================  
## Eigvenvalues with INTERCEPT  
## Intercept crim zn indus chas nox rm  
## Eigenvalues: 10.0928 1.5941 0.9599 0.6617 0.2411 0.1665 0.1084  
## Condition Indices: 1.0000 2.5162 3.2426 3.9054 6.4699 7.7856 9.6512  
## age dis rad tax ptratio black lstat  
## Eigenvalues: 0.0745 0.0416 0.0256 0.0132 0.0121 0.0072 0.0013  
## Condition Indices: 11.6402 15.5808 19.8404 27.6627 28.9124 37.4177 87.3183

# Plot of multicollinearity

mc.plot(x, y, Inter = FALSE, vif = 10, ev = 0.01)  
mc.plot(x, y)



mc.plot(x, y, vif = 5, ev = 0.2)



For Further Details about it re fer to this link

<http://rfaqs.com/mctest-r-package-detection-collinearity-among-regressors>