Package: rvif (via r-universe)

August 19, 2024

Type Package

Title Collinearity Detection using RVIF and Graphical Methods

Version 1.0

Date 2022-11-02

Author R. Salmeron and C. Garcia

Maintainer R. Salmeron <romansg@ugr.es>

Description The detection of troubling approximate collinearity in a
multiple linear regression model is a classical problem in
Econometrics. The objective of this package is to detect it
using the variance inflation factor redefined and the
scatterplot between the variance inflation factor and the
coefficient of variation.

License GPL (>= 2)

URL <http://colldetreat.r-forge.r-project.org/>

Depends R (>= 3.5.0), multiColl

Repository https://r-forge.r-universe.dev

RemoteUrl https://github.com/r-forge/colldetreat

RemoteRef HEAD

RemoteSha 4246eff3c5bf7ba7d28c2f283ad8833cf64ce286

Contents

rvif-package	*Multicollinearity Detection using RVIF and graphical methods*

Description

The detection of troubling near multicollinearity in a multiple linear regression model is a classical problem in Econometrics. The purpose of this package is its detection by using the Redefined Variance Inflation Factor (RVIF) and the scatterplot between the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV).

Details

This package contains two functions. On the one hand, CV_VIF, provides the values of the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV), as well as its representation in a scatter plot. Taking into account that the VIF is useful for detecting essential multicollinearity and the CV is useful for detecting non-essential multicollinearity, the scatter plot of both measures can provide interesting information for detecting whether there is a troubling degree of multicollinearity, what kind of multicollinearity it is and which variables are causing the multicollinearity.

On the other hand, RVIF, calculate the redefined VIF, the percentage of near multicollinearity due to each independent variable and, using the above function, the catter plot between the CV and VIF.

Author(s)

Román Salmerón Gómez (University of Granada) and Catalina García García (University of Granada).

Maintainer: Román Salmerón Gómez (romansg@ugr.es)

References

R. Salmerón, C. García, and J. García. Variance inflation factor and condition number in multiple linear regression. Journal of Statistical Computation and Simulation, 88:2365-2384, 2018.

R. Salmerón, A. Rodríguez, and C. García. Diagnosis and quantification of the non-essential collinearity. Computational Statistics, 35:647-666, 2020.

Limitations in Detecting Multicollinearity due to Scaling Issues in the mcvis Package by Salmerón, R., García, C.B, Rodríguez, A. and García, C. (working paper).

A redefined VIF by Salmerón, R., García, C.B, García, J. (working paper).

CV_VIF	*VIF, CV and its scatter plot*

Description

This function provides the values for the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV), as well as its representation in a scatter plot.

Usage

```
CV_VIF(X, size=NULL, top=82.64, limit=40, dummy=FALSE, pos=NULL, intercept=TRUE)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).
size	A numeric vector containing the percentage of multicollinearity due to each variable. By default size=NULL.
top	A real number that indicates the threshold from which the percentage of multi-collinearity due to each variable is considered troubling. By default top=82.64.
limit	A real number that indicates the lower limit of the vertical axis. By default limit=40.
dummy	A logical value that indicates if there are dummy variables in the design matrix X. By default dummy=FALSE.
pos	A numeric vector that indicates the position of the dummy variables, if these exist, in the design matrix X. By default pos=NULL.
intercept	A logical value used only by the function RVIF. By default intercept=TRUE.

Details

It is interesting to note the distinction between essential (near-linear relationship between at least two independent variables excluding the intercept) and non-essential multicollinearity (near-linear relationship between the intercept and at least one of the remaining independent variables), due to the VIF is not an appropriate measure to detect non-essential collinearity (only detects essential collinearity), while the CV is useful to detect only non-essential collinearity.

Then, this distinction between essential and non-essential multicollinearity and the limitations of each measure for detecting the different kinds of multicollinearity, can be very useful for detecting whether there is a troubling degree of multicollinearity, what kind of multicollinearity it is and which variables are causing the multicollinearity.

For this it is important include in the figures the lines corresponding to the established thresholds for each measure (CV and VIF): dashed vertical line for 0.1002506 (CV) and dotted horizontal line for 10 (VIF). These lines determine four regions (see Example 1) that can be interpreted as follows: A, existence of troubling non-essential and non-troubling essential multicollinearity; B, existence of troubling essential and non-essential multicollinearity; C, existence of non-troubling non-essential and troubling essential multicollinearity; D: non-troubling degree of existing multicollinearity (essential and non-essential).

Value

CV	Coefficient of Variation of each independent variable.
VIF	Variance Inflation Factor of each independent variable.

Author(s)

R. Salmerón (<romansg@ugr.es>) and C. García (<cbgarcia@ugr.es>).

References

R. Salmerón, C. García, and J. García. Variance inflation factor and condition number in multiple linear regression. Journal of Statistical Computation and Simulation, 88:2365-2384, 2018.

R. Salmerón, A. Rodríguez, and C. García. Diagnosis and quantification of the non-essential collinearity. Computational Statistics, 35:647-666, 2020.

Limitations in Detecting Multicollinearity due to Scaling Issues in the mcvis Package by Salmerón, R., García, C.B, Rodríguez, A. and García, C. (working paper).

Examples

```
## Example 1
plot(-2:20, -2:20, type = "n", xlab="Coefficient of Variation", ylab="Variance Inflation Factor")
abline(h=10, col="black", lwd=3, lty=2)
abline(v=0.1002506, col="black", lwd=3, lty=3)
text(-1.25, 2, "A", pos=3, col="red")
text(-1.25, 12, "B", pos=3, col="red")
text(10, 12, "C", pos=3, col="red")
text(10, 2, "D", pos=3, col="red")

## Example 2
library(multiColl)
set.seed(2022)
obs = 100
cte = rep(1, obs)
x2 = rnorm(obs, 5, 0.01)
x3 = rnorm(obs, 5, 10)
x4 = x3 + rnorm(obs, 5, 1)
x5 = rnorm(obs, -1, 30)
x = cbind(cte, x2, x3, x4, x5)
CV_VIF(x, size = c(1, 1, 1, 1))
```

RVIF	*RVIF calculation*

Description

This function provides the values of the Redefined Variance Inflation Factor (RVIF) and the the percentage of near multicollinearity due to each independent variable.

Usage

```
RVIF(X, l_u=TRUE, l=40, intercept=TRUE, graf=TRUE)
```

Arguments

X	A numeric design matrix that should contain more than one regressor.
l_u	A logical value that indicates if the variables in the design matrix X are transformed to have unit length. By default l_u=TRUE.
l	A real number that indicates the lower limit of the vertical axis of the scatter plot between the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV). By default l=40.
intercept	A logical value that indicates if the design matrix X have intercept. By default intercept=TRUE.
graf	A logical value that indicates if the scatter plot between the VIF and CV is represented by using CV_VIF function. By default graf=TRUE.

Details

The Redefined Variation Inflation Factor (RVIF) is able to detect both kind of multicollinearity: the essential (near-linear relationship between at least two independent variables excluding the intercept) an non-essential (near-linear relationship between the intercept and at least one of the remaining independent variables). This measure also quantifies the percentage of near multicollinearity due to each independent variable.

Value

RVIF	Redefined Variance Inflation Factor of each independent variable.
%	Percentage of near multicollinearity due to each independent variable.
Graph	Scatter plot of VIF and the CV.

Author(s)

R. Salmerón (<romansg@ugr.es>) and C. García (<cbgarcia@ugr.es>).

References

R. Salmerón, C. García, and J. García. Variance inflation factor and condition number in multiple linear regression. Journal of Statistical Computation and Simulation, 88:2365-2384, 2018.

R. Salmerón, A. Rodríguez, and C. García. Diagnosis and quantification of the non-essential collinearity. Computational Statistics, 35:647-666, 2020.

A redefined VIF by Salmerón, R., García, C.B, García, J. (working paper).

See Also

[CV_VIF](CV_VIF)

Examples

```
library(multiColl)
set.seed(2022)
obs = 100
cte = rep(1, obs)
x2 = rnorm(obs, 5, 0.01)
x3 = rnorm(obs, 5, 10)
x4 = x3 + rnorm(obs, 5, 1)
x5 = rnorm(obs, -1, 30)
x = cbind(cte, x2, x3, x4, x5)
RVIF(x)
```

Index