Package ‘mctest’

June 29, 2020

Type Package
Title Multicollinearity Diagnostic Measures
Version 1.3.1
Description Package computes popular and widely used multicollinearity diagnostic measures <doi:10.17576/jsm-2019-4809-26> and <doi:10.32614/RJ-2016-062>. Package also indicates which regressors may be the reason of collinearity among regressors.
License GPL (>= 2)
Encoding UTF-8
Depends R (>= 3.4), stats
Imports graphics
LazyData TRUE
BugReports http://rfaqs.com/contact
Note Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan.
NeedsCompilation no
Author Imdad Ullah Muhammad [aut, cre]
          (https://orcid.org/0000-0002-1315-491X),
          Aslam Muhammad [aut, ctb]
Maintainer Imdad Ullah Muhammad <mimdadasad@gmail.com>
Repository CRAN
Date/Publication 2020-06-29 19:00:03 UTC

R topics documented:

mctest-package .................................................. 2
eigprop .......................................................... 2
Hald ............................................................... 4
imcdiag .......................................................... 5
mc.plot .......................................................... 8
mctest ........................................................... 10
omcdiag .......................................................... 12
Description

R package for computing popular and widely used multicollinearity diagnostic measures.

Details

This package contains functions for computing overall and individual multicollinearity diagnostic measures. The overall multicollinearity diagnostic measures are Determinant of correlation matrix, R-squared from regression of all \( x \)’s on \( y \), Farrar and Glauber chi-square test for detecting the strength of collinearity over the complete set of regressors, Condition Index, Sum of reciprocal of Eigenvalues, Theil’s and Red indicator. The individual multicollinearity diagnostic measures are Klein’s rule, variance inflation factor (VIF), Tolerance (TOL), Corrected VIF (CVIF), Leamer’s method, \( F \) & \( R^2 \) relation, Farrar & Glauber F-test, and IND1 & IND2 indicators proposed by the author. The package also indicates which regressors may be the reason of collinearity among regressors. The VIF values and eigenvalues can also be plotted. Some other statistics such as correlation matrix, Eigenvalues and condition indexes are also available in the package.

For a complete list of functions, use `library(help="mctest")`.

Author(s)

Muhammad Imdad Ullah, Muhammad Aslam

eigprop

Eigenvalues and Variance Decomposition Proportion

Description

Computes eigenvalues, condition indices and variance decomposition proportions of \( X'X \) or its related correlation matrix \( R \) (see Belsley et al. (1980) <doi: 10.1007/BF00426854> ; Belsley, 1991; Kendall, 1957 and Silvey, 1969).

Usage

eigprop(mod, na.rm = TRUE, Inter = TRUE, prop = 0.5, ...)

Arguments

mod

A model object, not necessarily type `lm`

na.rm

Whether to remove missing observations.

Inter

Whether to include or exclude Intercept term, by default `Inter = FALSE`.

prop

variance proportion default threshold, `prop = 0.5`.

... Extra argument(s) if used will be ignored.
Details

The `eigprop` function can be used to detect the existence of multicollinearity among regressors. The function computes eigenvalues, condition indices and variance decomposition proportions of regression coefficients. To check the linear dependencies associated with the corresponding eigenvalue, the `eigprop` compares variance proportion with threshold value (default is 0.5) and displays the proportions greater than given threshold from each row and column, if any. If `Inter = TRUE`, eigenvalues, condition indices and variance proportions are computed without intercept term. A list object of class "eigp" is returned:

Value

The `eigprop` objects are:

- `ev`: A vector of eigenvalues. By default `Inter = TRUE`, eigenvalues are returned with intercept term included in the $X$ matrix.
- `ci`: A vector of condition indices. By default `Inter = TRUE`, condition indices are returned with intercept term included in the $X$ matrix.
- `call`: The matched call.
- `Inter`: logical, if `TRUE` (the default value) eigenvalues, condition indices and variance proportions are returned with intercept term included.
- `pi`: A matrix of variance decomposition proportions. By default `Inter = TRUE`, variance decomposition proportions are returned with intercept term included in the $X$ matrix.
- `prop`: Default threshold proportion for comparison purpose.

Note

Missing values in data will be removed by default. There is no method for the detection of multicollinearity, if missing values exists in the data set.

Author(s)

Muhammad Imdad Ullah, Muhammad Aslam

References


See Also

Overall collinearity diagnostics `omcdiag`, Individual collinearity diagnostics `imcdiag`

Examples

```r
## Hald Cement data
data(Hald)
model <- lm(y~X1+X2+X3+X4, data = as.data.frame(Hald))
# with Intercept term
eigprop(model)
# without Intercept term
eigprop(model, Inter = FALSE)
# different proportion threshold
eigprop(model, prop = 0.45)
# only variance proportions
eigprop(model)$pi
# only condition indices
eigprop(model)$ci
# only eigenvalues
eigprop(model)$ev
```

Hald

Portland Cement benchmark of Hald(1952)

Description

Heat evolved during setting of 13 cement mixtures of four basic ingredients. Each ingredient percentage appears to be rounded down to a full integer. The sum of the four mixture percentages varies from a maximum of 99% to a minimum of 95%. If all four regressor X-variables always summed to 100%, the centered X-matrix would then be of rank only 3. Thus, the regression of heat on four X-percents is ill-conditioned, with an approximate rank deficiency of MCAL = 1. The first column is the response and the remaining four columns are the predictors.

The Hald data as used by Hoerl, Kennard and Baldwin (1975). These data are also in package `wle`.

Usage

data(Hald)
Format

A data frame with 13 observations on the following 5 variables.

Y  Heat (cals/gm) evolved in setting, recorded to nearest tenth.
X1 Integer percentage of 3CaO.Al2O3 in the mixture.
X2 Integer percentage of 3CaO.SiO2 in the mixture.
X3 Integer percentage of 4CaO.Al2O3.Fe2O3 in the mixture.
X4 Integer percentage of 2CaO.SiO2 in the mixture.

Source


References


Examples

data(Hald)
y <- Hald[, 1]
x <- Hald[, -1]

imcdiag

Individual Multicollinearity Diagnostic Measures

Description


Usage

imcdiag(mod, method = NULL, na.rm = TRUE, corr = FALSE,
         vif = 10, tol = 0.1, conf = 0.95, cvif = 10, ind1 = 0.02,
         ind2 = 0.7, leamer = 0.1, all = FALSE, ...)
Arguments

- `mod`: A model object, not necessarily type `lm`.
- `na.rm`: Whether to remove missing observations.
- `method`: Specific individual measure of collinearity such as VIF, CVIF, and Leamer, etc. For example, `method="VIF"`.
- `corr`: Whether to display correlation matrix or not, by default `corr=FALSE`.
- `vif`: Default threshold for VIF measure, `vif=10`.
- `tol`: Default threshold for TOL measure, `tol=0.10`.
- `conf`: Default confidence level for Farrar’s Wi test, `conf=0.99`.
- `cvif`: Default threshold for CVIF measure, `CVIF=10`.
- `ind1`: Default threshold for IND1 indicator, `ind1=0.02`.
- `ind2`: Default threshold for IND2 indicator, `ind2=0.7`.
- `leamer`: Default threshold for Leamer’s method, `leamer=0.1`.
- `all`: Returns all individual measure of collinearity in a matrix of 0 (not detected) or 1 (detected).
  - `...`: Extra argument(s) if used will be ignored.

Details

The `imcdiag` function detects the existence of multicollinearity due to `x`-variable. That’s why named as individual measures of diagnostics. This includes VIF, TOL, Klein’s rule, Farrar and Glauber F-test, F and $R^2$ relation, Leamer’s method, CVIF, IND1, and IND2 diagnostic measures of multicollinearity. If `method` argument is used (`method="VIF"`), the VIF values for each regressor will be displayed with decision of either collinearity exists or not which is indicated by 0 (collinearity is not detected by method for regressor) and 1 (collinearity is detected by the method for regressor). If argument `all=TRUE` all individual measures of collinearity will be displayed in a matrix of 0 (collinearity is not detected) or 1 (collinearity is detected).

Value

This function detects the existence of multicollinearity by using different available diagnostic measures already available in literature. The function returns the value of diagnostic measures with decision of either collinearity is detected by the diagnostic measure or not. Value of 1 indicates that collinearity is detected and 0 indicates that measure could not detect the existence of collinearity. A list object of class “imc” is returned:

- `x`: A numeric matrix of regressors.
- `y`: A vector of response variable.
- `idiags`: Listing of specific individual measure such as `method="CVIF"` provided. If method is not used all individual diagnostics will be displayed.
- `method`: Specific individual collinearity measure, such as VIF, TOL, CVIF, IND1, and IND2 etc.
- `corr`: Logical, if FALSE (the default value) a correlation matrix will not be displayed.
imcdiag

\textbf{R2} \hspace{1cm} \text{R-square from regression of all regressors $X$ on response variable $y$.}

\textbf{call} \hspace{1cm} \text{The matched call.}

\textbf{pval} \hspace{1cm} \text{Returns significant regressor as number after comparing the $p$-value of regressors from \texttt{summary.lm} function with $1 - \text{conf}$.}

\textbf{all} \hspace{1cm} \text{If \texttt{TRUE} individual collinearity measures will be returned as a matrix of 0 or 1.}

\textbf{alldiag} \hspace{1cm} \text{Matrix of all individual collinearity measures indicated as either 0 (collinearity not detected) or 1 (collinearity detected) for each diagnostic measure and each regressor.}

\textbf{Note}

Missing values in data will be removed by default. There is no method for the detection of multicollinearity, if missing values exists in the data set.

\textbf{Author(s)}

Muhammad Imdad Ullah, Muhammad Aslam

\textbf{References}


Imdad, M. U. \textit{Addressing Linear Regression Models with Correlated Regressors: Some Package Development in R} (Doctoral Thesis, Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan), 2017.


\textbf{See Also}

Overall collinearity diagnostic \texttt{omcdiag}, collinearity plot \texttt{mc.plot}
Examples

```r
## Hald Cement data
data(Hald)
model <- lm(y~X1+X2+X3+X4, data = as.data.frame(Hald))

## all Individual measures
id<-imcdiag(model); id$idiags[,1]

# VIF measure with custom VIF threshold
imcdiag(model, method = "VIF", vif = 5)

# IND1 measure with custom IND1 threshold and correlation matrix
imcdiag(model, method="IND1", ind1=0.01, corr=TRUE)

# CVIF measure with custom CVIF threshold and correlation matrix
imcdiag(model, method = "CVIF", cvif = 5, corr = TRUE)

# Collinearity Diagnostic measures in matrix of 0 or 1
imcdiag(model, all = TRUE)
imcdiag(model, method = "VIF", all = TRUE)

## only VIF values without collinearity detection indication
imcdiag(model, method = "VIF")[[1]][,1]
plot(imcdiag(model, method = "VIF")[[1]][,1]) # vif plot
```

Description

Plot of VIF and Eigenvalues for detection of multicollinearity among regressors. The VIF and Eigenvalues are also displayed on graph. Eigenvalues plot can be displayed with or without inclusion of intercept term.

Usage

```r
mc.plot(mod, Inter = FALSE, vif = 10, ev = 0.01, ...)
```

Arguments

- `mod`: A model object, not necessarily type `lm`
- `Inter`: Whether to include or exclude Intercept term, by default `Inter=FALSE`.
- `vif`: Threshold of VIF and will appear as horizontal line on VIF plot. The default value is `vif=10`.
- `ev`: Threshold of Eigenvalues and will appear as horizontal line on Eigenvalues plot. The default value is `ev=0.01`.
- `...`: Extra argument(s) if used will be ignored.
mc.plot

Details

mc.plot function draw graphs of VIF and Eigenvalues for graphical detection of collinearity among regression. Horizontal line for VIF and Eigenvalues is drawn as indication of threshold values of both VIF and Eigenvalues for testing the multicollinearity.

Value

Don’t return any thing, it displays plot.

Author(s)

Muhammad Imdad Ullah, Muhammad Aslam

References


See Also

Overall collinearity diagnostic measures omcdiag, Individual collinearity diagnostic measures imcdiag

Examples

```r
## Hald Cement data
data(Hald)
model <- lm(y~X1+X2+X3+X4, data = as.data.frame(Hald))

## plot with default threshold of VIF and Eigenvalues with no intercept
mc.plot(model)

## plot with default threshold of VIF and Eigenvalues with intercept
mc.plot(model, Inter = TRUE)

## plot with specific threshold of VIF and Eigenvalues with no intercept
mc.plot(model, vif = 5, ev = 20)
```
### Multicollinearity diagnostic measures

**Description**

The function `mctest` display overall, individual or both types of multicollinearity diagnostic measures from `omcdiag` and `imcdiag` functions, respectively.

**Usage**

```r
mctest(mod, type=c("o","i","b"), na.rm=TRUE, Inter=TRUE, method=NULL, 
corr=FALSE, detr=0.01, red=0.5, theil=0.5, cn=30, vif=10, tol=0.1, 
conf=0.95, cvif=10, ind1=0.02, ind2=0.7, leamer=0.1, all=FALSE, ...)
```

**Arguments**

- `mod`: A model object, not necessarily type `lm`.
- `na.rm`: Whether to remove missing observations.
- `Inter`: Whether to include or exclude Intercept term. By default `Inter=TRUE`.
- `type`: Displays overall, individual or both type of collinearity diagnostics. Overall collinearity diagnostics are displayed by default with eigenvalues and condition indexes, when method and type argument are not used.
- `method`: Specific individual measure of collinearity such as VIF, TOL, CVIF, Leamer, IND1, and IND2 etc, when method argument is used. For example, `method="VIF"`.
- `corr`: Whether to display correlation matrix or not `Inter=TRUE`.
- `detr`: Determinant default threshold, `detr=0.01`.
- `red`: Red indicator default threshold, `red=0.5`.
- `theil`: Theil’s indicator default threshold, `theil=0.5`.
- `cn`: Condition number default threshold, `cn=30`.
- `vif`: Default threshold for VIF measure, `vif=10`.
- `conf`: Default confidence level for Farrar’s test, `conf=0.99`.
- `cvif`: Default threshold for CVIF measure, `CVIF=10`.
- `tol`: Default threshold for TOL measure, `TOL=0.10`.
- `ind1`: Default threshold for IND1 indicator, `ind1=0.02`.
- `ind2`: Default threshold for IND2 indicator, `ind2=0.7`.
- `leamer`: Default threshold for Leamer’s method, `leamer=0.1`.
- `all`: Returns all individual measure of collinearity in a matrix of 0 (not detected) or 1 (detected).
- `...`: Extra argument(s) if used will be ignored.
Note

Missing values in data will be removed by default. There is no method for the detection of multicollinearity, if missing values exists in the data set

Author(s)

Muhammad Imdad Ullah, Muhammad Aslam

References


See Also

overall collinearity diagnostics [omcdiag](https://example.com), individual collinearity diagnostics [imcdiag](https://example.com), collinearity plots [mc.plot](https://example.com)

Examples

```r
## Hald Cement data
data(Hald)
model <- lm(y~X1+X2+X3+X4, data = as.data.frame(Hald))

## Overall diagnostic measures and eigenvalues with intercept term
mctest(model)

## Overall diagnostic measures and eigenvalues without intercept term
mctest(model, Inter=FALSE)

## all individual diagnostic measures
mctest(model, type="i")

## certain individual diagnostic measures with collinearity detection indication
VIF<-mctest(model, type="i", method="VIF")
```
VIF[[1]][,1] # named VIF values only

IND1<-mctest(model, type="i", method="IND1")
IND1[[1]][,1] # named IND1 values only

## all individual diagnostic measures with correlation matrix
mctest(model, type="i", corr=TRUE)

## VIF and correlation matrix with collinearity detection indication
mctest(model, type="i", method="VIF", corr=TRUE)

## both overall and individual collinearity diagnostics
mctest(model, type="b")

## VIF and CN desired threshold
eigenvalues without intercept term
mctest(model, type="b", method="VIF", Inter=FALSE, vif=15, cn=35)

## Individual collinearity diagnostic measures in matrix of 0 or 1
mctest(model, all = TRUE)
mctest(model, method = "VIF", all = TRUE)

omcdiag

Overall Multicollinearity Diagnostics Measures

Description

Usage

omcdiag(mod, na.rm = TRUE, Inter = TRUE, detr = 0.01, red = 0.5, conf = 0.95, theil = 0.5, cn = 30,...)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mod</td>
<td>A model object, not necessarily type lm</td>
</tr>
<tr>
<td>na.rm</td>
<td>Whether to remove missing observations.</td>
</tr>
<tr>
<td>Inter</td>
<td>Whether to include or exclude Intercept term, by default Inter=TRUE.</td>
</tr>
</tbody>
</table>
omcdiag

\begin{itemize}
\item \texttt{detr} \quad \text{Determinant default threshold, detr} = 0.01.
\item \texttt{red} \quad \text{red indicator default threshold, red} = 0.5.
\item \texttt{conf} \quad \text{confidence level of Farrar Chi-Square test, conf} = 0.95.
\item \texttt{theil} \quad \text{Theil's indicator default threshold, theil} = 0.5.
\item \texttt{cn} \quad \text{condition number default threshold, cn} = 30.
\item \ldots \quad \text{Extra argument(s) if used will be ignored.}
\end{itemize}

**Details**

This function detects the existence of multicollinearity by using different available diagnostic measures already available in literature such as Determinant of correlation matrix, Farrar test of chi-square, Red Indicator, Sum of lambda inverse values, Theil's Indicator and Condition Number.

Function also displays diagnostic measures value with the decision of either multicollinearity is detected by the diagnostics or not. The Value of 1 indicate that multicollinearity is detected and 0 indicate measure could not detect by the certain diagnostic measure. A list object of class "omc" is returned:

**Value**

\begin{itemize}
\item \texttt{odiags} \quad \text{Listing of all overall diagnostic measures.}
\item \texttt{inter} \quad \text{logical, if TRUE (the default value) condition number is returned with intercept term included.}
\item \texttt{x} \quad \text{matrix of regressors.}
\item \texttt{call} \quad \text{The matched call.}
\end{itemize}

**Note**

Missing values in data will be removed by default. There is no method for the detection of multicollinearity, if missing values exists in the data set.

**Author(s)**

Muhammad Imdad Ullah, Muhammad Aslam

**References**


### See Also

Individual collinearity diagnostic measure `imcdiag`, Eigenvalues and variance decomposition proportion `eigprop`

### Examples

```R
## Hald Cement data
data(Hald)
model <- lm(y~X1+X2+X3+X4, data = as.data.frame(Hald))

## all overall diagnostic measures and eigenvalues with intercept
od<-omcdiag(model)

## all overall diagnostic measures and eigenvalues without intercept
omcdiag(model, Inter=FALSE)

## all overall diagnostic measures and eigenvalues with intercept
## with different determinant and confidence level threshold
omcdiag(model, detr=0.001, conf=0.99)

## returns the determinant of correlation matrix |X'X|
omcdiag(model)[1]
```
Index

* CVIF
  imcdiag, 5
* Collinearity Diagnostics Plots
  mc.plot, 8
* Condition Number
  omcdiag, 12
* Determinant of correlation matrix
  omcdiag, 12
* Eigenvalues Plot
  mc.plot, 8
* F-test and R-Square relation
  imcdiag, 5
* Farrar Chi-Square
  omcdiag, 12
* Farrar F-test
  imcdiag, 5
* IND1
  imcdiag, 5
* IND2
  imcdiag, 5
* Individual collinearity diagnostic measures
  mctest, 10
* Leamer's Method
  imcdiag, 5
* Multicollinearity Detection
  imcdiag, 5
* Multicollinearity Diagnostics
  mctest, 10
  mctest-package, 2
* Overall Multicollinearity Diagnostics
  omcdiag, 12
* Overall collinearity diagnostic measures
  mctest, 10
* Red Indicator
  omcdiag, 12
* Regression Diagnostics
  mctest-package, 2
* TOL
  imcdiag, 5
* Theil's Indicator
  omcdiag, 12
* VIF Plot
  mc.plot, 8
* VIF
  imcdiag, 5
* condition indices
  eigprop, 2
* condition number
  eigprop, 2
* datasets
  Hald, 4
* eigenvalues
  eigprop, 2
* klein's Rule
  imcdiag, 5
* mctest package
  mctest-package, 2
* multicollinearity
  eigprop, 2
* sum of lambda inverse values
  omcdiag, 12
* variance decomposition proportion
  eigprop, 2

Hald, 4

imcdiag, 4, 5, 9, 11, 14
mc.plot, 7, 8, 11
mctest, 10
mctest-package, 2
omcdiag, 4, 7, 9, 11, 12
print.eigp(eigprop), 2
print.imc(imcdiag), 5
print.omc(omcdiag), 12