Package 'multiColl'

July 18, 2019

Type Package

Title Collinearity Detection in a Multiple Linear Regression Model

Version 1.0

Date 2019-07-07

Author R. Salmeron, C.B. Garcia and J. Garcia

Maintainer R. Salmeron <romansg@ugr.es>

Description The detection of worrying approximate collinearity in a multiple linear regression model is a problem addressed in all existing statistical packages. However, we have detected deficits regarding to the incorrect treatment of qualitative independent variables and the role of the intercept of the model. The objective of this package is to correct these deficits. In this package will be available detection and treatment techniques traditionally used as the recently developed.
D.A. Belsley (1982) <doi:10.1016/0304-4076(82)90020-3>.
D. A. Belsley (1991, ISBN: 978-0471528890).
C. Garcia, R. Salmeron and C.B. Garcia (2019) <doi:10.1080/00949655.2018.1543423>.
R. Salmeron, C.B. Garcia and J. Garcia (2018) <doi:10.1080/00949655.2018.1463376>.
G.W. Stewart (1987) <doi:10.1214/ss/1177013444>.

License GPL (>= 2)

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

Repository CRAN

Repository/R-Forge/Project colldetreat

Repository/R-Forge/Revision 26

Repository/R-Forge/DateTimeStamp 2019-07-15 18:37:09

Date/Publication 2019-07-18 11:14:05 UTC

NeedsCompilation no

R topics documented:

`multiColl-package` *Collinearity detection in a multiple linear regression model.*

Description

R package to detect collinearity in a multiple linear regression model.

Details

The detection of worrying approximate multicollinearity in a multiple linear regression model is a problem addressed in all existing statistical packages. However, we have detected deficits regarding to the incorrect treatment of qualitative independent variables and the role of the intercept of the model. The objective of this package is to correct these deficits. In this package will be available detection and treatment techniques traditionally used as the recently developed.

Author(s)

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

References

`multiColl`: An R Package for Detecting Multicollinearity. Working paper.

CN *Condition Number*

Description

This function returns the Condition Number (CN) of the independent variables in a multiple linear regression.

Usage

```
CN(X)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).

Details

Due to the CN takes into account the intercept, it allows to detect not only the essential but also the non-essential collinearity. It also allows to consider non-quantitative independent variables.

Its calculation is obtained from the function `lu`, contrary to the function `kappa`.

Value

The condition number of a matrix, that is, the maximum condition index.

Note

Values of CN between 20 and 30 indicate near moderate multicollinearity while values higher than 30 indicate near worrying collinearity.

Author(s)

R. Salmeron (<romansg@ugr.es>) and C. Garcia (<cbgarcia@ugr.es>).

References

D. A. Belsley (1991). Conditioning diagnostics: collinearity and weak dara in regression. John Wiley & Sons, New York.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

`lu`, `kappa`, `CNs`.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
CN(theil.X)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
CN(KG.X)
```

CNs *Condition Number with and without intercept*

Description

This function returns the Condition Number (CN) of the independent variables of a multiple linear model considering the intercept and without considering it. It also returns the increase produced by going from not taking into account the intercept to having it.

Usage

```
CNs(X)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).

Value

CN1	Condition Number without intercept.
CN2	Condition Number with intercept.
increment	Increase (in percentage) in the CN from CN1 to CN2.

Author(s)

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

References

D. A. Belsley (1991). Conditioning diagnostics: collinearity and weak data in regression. John Wiley & Sons, New York.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

`lu`, `CN`.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
CNs(theil.X)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
CNs(KG.X)
```

CV *Coeficient of Variation*

Description

The function calculates the Coeficient of Variation (CV) of a quantitative vector.

Usage

```
CV(x)
```

Arguments

`x`	A quantitative vector.

Value

The CV of `x`.

Author(s)

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

See Also

mean, var, sd.

Examples

```
# random
x = sample(1:50, 25)
x
CV(x)
```

CVs *Coeficients of Variation*

Description

The function returns the Coeficient of Variation (CV) of a matrix with quantitative columns.

Usage

```
CVs(X, dummy = FALSE, pos = NULL)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).
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.

Details

Due to the calculation of the CV only makes sense for quantitative data, other kind of data should be ignored in the calculation. For this reason, it is necessary to indicate if there are non-quantitative variables and also its position in the matrix.

Value

The CV of each column of X.

Author(s)

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

See Also

CV.

Examples

```
# random

cte = array(1, 50)
x1 = sample(1:50, 25)
x2 = sample(1:50, 25)
Z = cbind(cte, x1, x2)
head(Z)
CVs(Z)

x3 = sample(c(array(1,25), array(0,25)), 25)
W = cbind(Z, x3)
head(W)
CVs(W, dummy=TRUE, pos = 4)

x0 = sample(c(array(1,25), array(0,25)), 25)
Y = cbind(cte, x0, x1, x2, x3)
head(Y)
CVs(Y, dummy=TRUE, pos=c(2,5))
```

KG *Klein and Goldberger data*

Description

Klein and Goldberger data on consumption and wage income.

Usage

```
data("KG")
```

Format

A data frame with 14 observations in relation to the following four variables:

`consumption` Domestic consumption.

`wage.income` Wage income.

`non.farm.income` Non-wage-non-farm income.

`farm.income` Farm income.

Details

Data for the years 1942 to 1944 are not available for the war.

References

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

Examples

```
data(KG)
head(KG)
```

`ki` *Stewart's index*

Description

The function returns the index of Stewart of the independent variables in the multiple linear regession model.

Usage

```
ki(X, dummy = FALSE, pos = NULL)
```

Arguments

`X`	A numeric design matrix that should contain more than one regressor (intercept included).
`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`.

Details

The index of Stewart allows to detect the near essential and non-essential multicollinearity existing in a multiple linear regression model. In addition, due to its relation with the Variance Inflation Factor (VIF), it allows to calculate the proportion of essential and non-essential multicollinearity in each independent variable (intercept excluded). The Stewart's index for the intercept indicates the degree of non-essential multicollinearity existing in the model.

The relation of the the VIF with the index of Stewart implies that it should not be calculated for non-quantitative variables.

Value

`ki`	Stewart's index for each independent variable.
`porc1`	Proportion of essential multicollinearity in the i-th independent variable (without intercept).
`porc2`	Proportion of non-essential multicollinearity in the i-th independent variable (without intercept).

Author(s)

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

References

G. Stewart (1987). Collinearity and least squares regression. Statistical Science, 2 (1), 68-100.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

VIF.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
ki(theil.X, TRUE, pos = 4)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
ki(KG.X)
```

---

lu *Unit length data*

Description

The function transforms the matrix X so that each column has unit length, it is to say, a module equal to 1.

Usage

```
lu(X)
```

Arguments

X	A numeric matrix that should contain more than one column.

Value

Original matrix transformed so that each column has a module equal to 1.

Author(s)

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

References

R. Salmerón, C. B. García and J. García (2018). Variance Inflation Factor and Condition Number in multiple linear regression. Journal of Statistical Computation and Simulation, 88 (12), 2365-2384.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

`CN`, `CNs`.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
lu(theil.X)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
lu(KG.X)

# random
x1 = sample(1:10,5)
x2 = sample(1:10,5)
x = cbind(x1, x2)
x
norm(x[,1],"2")
norm(x[,2],"2")
x.lu = lu(x)
x.lu
norm(x.lu[,1],"2")
norm(x.lu[,2],"2")
```

`multiCol` *Collinearity detection in a linear regression model*

Description

The function collects all existing measures to detect worrying multicollinearity in the package `multiCol`.

Usage

```
multiCol(X, dummy = FALSE, pos = NULL)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).
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.

Value

If X contains two independent variables (intercept included) see SLM function.

If X contains more than two independent variables (intercept included):

CV	Coeficients of variation of quantitative variables in X.
Prop	Proportion of ones in the dummy variables.
R	Matrix correlation of the quantitative variables in X.
detR	Determinant of the matrix correlation of the quantitative variables in X.
VIF	Variance Inflation Factors of the quantitative variables in X.
CN	Condition Number of X.
ki	Stewart's index of the quantitative variables in X.

Note

For more detail, see the help of the functions in See Also.

Author(s)

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

References

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

SLM, CV, PROPs, RdetR, VIF, CN, ki.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
multiCol(theil.X, TRUE, pos = 4)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
multiCol(KG.X)

# random
x1 = array(1,25)
x2 = rnorm(25,100,1)
x = cbind(x1,x2)
head(x)
multiCol(x)

# random
x1 = array(1,25)
x2 = sample(cbind(array(1,25),array(0,25)),25)
x = cbind(x1,x2)
head(x)
multiCol(x, TRUE)
```

`multiColLM` *All detection measures*

Description

The functions collects all the measure to detect near worrying multicollinearity existing in the package `multiCol`. In adddition, it provides the estimations by ordinary least squares (OLS) of the multiple linear regession model and the variations in the estimations of the coefficients as a consequence of changes in the observed data.

Usage

```
multiColLM(y, X, dummy = FALSE, pos1 = NULL, n, mu, dv, tol = 0.01, pos2 = NULL)
```

Arguments

`y`	Observations of the dependent variable of the model.
`X`	Observations of the independent variables of the model (intercept included).
`dummy`	A logical value that indicates if there are dummy variables in the design matrix `X`. By default `dummy=FALSE`.

pos1	A numeric vector that indicates the position of the dummy variables, if these exist, in the design matrix X. By default pos=NULL.
n	Number of times that the perturbation is performed.
mu	Any real number.
dv	Any real positive number.
tol	A value between 0 and 1. By default tol=0.01.
pos2	A numeric vector that indicates the position of the independent variables to disturb once you eliminate in data the dependent variable and the intercept. By default pos=NULL.

Value

The estimation by OLS of the linear regression model.

Percentiles 2.5 and 97.5 of the proportion of the variations in the estimations of the coefficients obtained from a perturbation of tol% in the quantitative variables of X.

If X contains two independent variables (intercept included) see SLM function.

If X contains more than two independent variables (intercept included):

CV	Coeficients of variation of quantitative variables in X.
Prop	Proportion of ones in the dummy variables.
R	Matrix correlation of the quantitative variables in X.
detR	Determinant of the matrix correlation of the quantitative variables in X.
VIF	Variance Inflation Factors of the quantitative variables in X.
CN	Condition Number of X.
ki	Stewart's index of the quantitative variables in X.

Note

For more detail, see the help of the functions in See Also.

Author(s)

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

References

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

SLM, CV, PROPs, RdetR, VIF, CN, ki, multiCol, perturb, perturb.n.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
head(theil.X)
multiColLM(theil[,2], theil.X, dummy = TRUE, pos1 = 4, 5, 5, 5, tol=0.01, pos2 = 1:2)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
head(KG.X)
multiColLM(KG[,1], KG.X, n = 500, mu = 5, dv = 5, tol=0.01, pos2 = 1:3)
```

`perturb` *Perturbation*

Description

The function modifies a set of quantitative data.

Usage

```
perturb(x, mu, dv, tol = 0.01)
```

Arguments

`x`	A numeric quantitative vector.
`mu`	Any real number.
`dv`	Any real positive number.
`tol`	A value between 0 and 1. By default `tol=0.01`.

Details

The vector of data set is modified a `tol%` by following the procedure presented by Belsley (1982).

Value

The vector `x` modified a `tol%`.

Author(s)

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

References

D. Belsley (1982). Assessing the presence of harmfull collinearity and other forms of weak data throught a test for signal-to-noise. Journal of Econometrics, 20, 211-253.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

`perturb.n`.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
consume.p1 = perturb(theil[,2], 3, 4, 0.01)
consume.p2 = perturb(theil[,2], 50, 10, 0.01)
x = cbind(theil[,2], consume.p1, consume.p2)
head(x)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
farm.income.p1 = perturb(KG[,4], -3, 40, 0.01)
farm.income.p2 = perturb(KG[,4], 10, 8, 0.01)
x = cbind(KG[,4], farm.income.p1, farm.income.p2)
head(x)
```

`perturb.n` *Perturbation and estimation in a multiple linear model*

Description

The function quantifies the variations in the estimations of the coefficients of a multiple linear regression when a perturbation is introduced in the quantitative data set.

Usage

```
perturb.n(data, n, mu, dv, tol = 0.01, pos = NULL)
```

Arguments

`data`	Data set `(y,X)` where `y` and `X` contain, respectively, the observations of the dependent variable and independients variables (intercept included) of the multiple linear regression.
`n`	Number of times that perturbation is performed.

mu	Any real number.
dv	Any real positive number.
tol	A value between 0 and 1. By default tol=0.01.
pos	A numeric vector that indicates the position of the independent variables to disturb once you eliminate in data the dependent variable and the intercept. By default pos=NULL.

Value

tols	A vector presenting the percentage of disturbance induced in the variables indicated in each iteration.
norms	A vector presenting the percentage of variation in the estimations of the coefficients in each iteration.

Note

tols must be a constant vector equal to tol. It is obtained to check if data have been correctly perturbed.

Author(s)

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

References

D. Belsley (1982). Assessing the presence of harmfull collinearity and other forms of weak data throught a test for signal-to-noise. Journal of Econometrics, 20, 211-253.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

perturb.

Examples

```
tol = 0.01
mu = 10
dv = 10

# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.y.X = cbind(theil[,2], cte, theil[,-(1:2)])
head(theil.y.X)

iterations = 5
```

```
perturb.n.T = perturb.n(theil.y.X, iterations, mu, dv, tol, pos = c(1,2))
perturb.n.T
mean(perturb.n.T[,1])
mean(perturb.n.T[,2])
c(min(perturb.n.T[,2]), max(perturb.n.T[,2]))

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.y.X = cbind(KG[,1], cte, KG[,-1])
head(KG.y.X)

iterations = 1000

perturb.n.KG = perturb.n(KG.y.X, iterations, mu, dv, tol, pos = c(1,2,3))
mean(perturb.n.KG[,1])
mean(perturb.n.KG[,2])
c(min(perturb.n.KG[,2]), max(perturb.n.KG[,2]))
```

`PROPs` *Proportions*

Description

The functions returns the proportion of ones in the dummy variables existing in a matrix.

Usage

```
PROPs(X, dummy = TRUE, pos = NULL)
```

Arguments

`X`	A numeric matrix that should contain more than one regressor (intercept included).
`dummy`	A logical value that indicates if there are dummy variables in the matrix `X`. By default `dummy=TRUE`.
`pos`	A numeric vector that indicates the position of the dummy variables, if these exist, in the matrix `X`. By default `pos=NULL`.

Author(s)

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

See Also

`multiCol`.

Examples

```
# random
x1 = sample(1:50, 25)
x2 = sample(1:50, 25)
x3 = sample(cbind(array(1,25), array(0,25)), 25)
x4 = sample(cbind(array(1,25), array(0,25)), 25)
x = cbind(x1, x2, x3, x4)
head(x)
PROPs(x, TRUE, pos = c(3,4))
```

RdetR *Correlation matrix and it's determinat*

Description

The function returns the matrix of simple linear correlations between the independent variables of a multiple linear model and its determinant.

Usage

```
RdetR(X, dummy = FALSE, pos = NULL)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).
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.

Details

The measures calculated by this function ignore completelly the role of the intercept in the linear relations between the independent variables. Thus, these measures only detect the near essential multicollinearity. Although the simple correlations only quantify relation between pairs of variables, the determinant of the matrix of correlations is able to detect broader relations. Due to the coefficients of simple linear regression are calculated for quantitative variables, if the model contains other kinds of variables (such as dummy variables), they should be omitted in the analysis by using the arguments dummy and pos.

Value

R	Correlation matrix of the independent variables of the multiple linear regression model.
detR	Determinant of R.

Note

Values of the coefficient of simple linear correlation higher than 0.9487 imply worrying near essential multicollinearity between pairs of variables.

Values of the determinant of `R` lower than `0.1013 + 0.00008626*n -0.01384*k`, where `n` is the number of observations and `k` the number of indepedent variables (intercept included), indicate worrying near essential multicollinearity.

Author(s)

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

References

C. García, R. Salmerón and C. B. García (2019). Choice of the ridge factor from the correlation matrix determinant. Journal of Statistical Computation and Simulation, 89 (2), 211-231.

D. Marquardt and R. Snee (1975). Ridge regression in practice. The American Statistician, 1 (29), 3–20.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

`VIF`.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
RdetR(theil.X, TRUE, pos = 4)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
RdetR(KG.X)
```

SLM *Simple linear regression model and multicollinearity*

Description

The function analyzes the presence of near worrying multicollinearity in the Simple Linear Model (SLM).

Usage

```
SLM(X, dummy = FALSE)
```

Arguments

X	A numeric design matrix that should contain two independent variables (intercept included).
dummy	A logical value that indicates if there are dummy variables in the design matrix X. By default dummy=FALSE.

Details

The analysis of the presence of near worrying multicolllinearity in the SLM has been systematically ignored in some existing statistical softwares. However, it is possible to find worrying non essential multicollinearity in the SLM. In this case, the linear relation will be given by a second variable of X with very little variablity. For this reason, the coeficient of variation is calculated when the variable is quantitative and the proportion of ones if the variable is non-quantitative.

Value

If dummy=TRUE:

Prop	Proportion of ones in the dummy variable.
CN	Condition Number of X.

If dummy=FALSE:

CV	Coeficient of variation of the second variable in X.
VIF	Variance Inflation Factor.
CN	Condition Number of X.
ki	Stewart's index of X.

Note

The VIF only detects the near essential multicollinearity and for this reason it is not appropriate to detect multicollinearity in the SLM. Indeed, in this case, the VIF will be always equal to 1.

Author(s)

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

References

R. Salmerón, C. B. García and J. García (2018). Variance Inflation Factor and Condition Number in multiple linear regression. Journal of Statistical Computation and Simulation, 88 (12), 2365-2384.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

PROPs, CV, CN, ki.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
SLM(theil.X, TRUE)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
SLM(KG.X)

# random
x1 = array(1,25)
x2 = sample(1:50,25)
x = cbind(x1,x2)
head(x)
SLM(x)

# random
x1 = array(1,25)
x2 = rnorm(25,100,1)
x = cbind(x1,x2)
head(x)
SLM(x)

# random
x1 = array(1,25)
x2 = sample(cbind(array(1,25),array(0,25)),25)
x = cbind(x1,x2)
head(x)
SLM(x, TRUE)
```

`theil` *Henri Theil data*

Description

Henri Theil's textile consumption data modified.

Usage

```
data("theil")
```

Format

A data set with 17 observations in relation to the following five variables:

`obs` Year.

`consume` Volume of textile consumption per capita (base 1925=100).

`income` Real Income per capita (base 1925=100).

`relprice` Relative price of textiles (base 1925=100).

`twentys` Dummy variable that differentiates between the twenties and thirties.

Details

This data set is developed based on the original Henri Theil's textile consumption data. With the goal of showing the treatment of the detection of collinearity when non-quantitative variables exists in the multiple linear regression, a new dummy variable has been incorporated distinguishing between the twenties and thirties.

References

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

Examples

```
data(theil)
head(theil)
```

VIF *Variance Inflation Factor*

Description

The function returns the Variance Inflation Factors (VIFs) of the independent variables of the multiple linear regression model.

Usage

```
VIF(X, dummy = FALSE, pos = NULL)
```

Arguments

X	A numeric design matrix that should contain more than one regressor (intercept included).
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.

Details

The function returns the VIFs from the main diagonal of the inverse of the matrix of correlations of the independent variables of the multiple linear regression. Due to the VIF is only calculated for the independent variables, it only allows to detect the essential collinearity. In addition, the VIF is not adequate for dummy variables since it is obtained from the matrix of simple correlations.

Value

Variance Inflation Factor of each independent variable excluded the intercept.

Note

Values of VIF that exceed 10 indicate near essential multicolinearity.

Author(s)

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

References

D. Marquardt and R. Snee (1975). Ridge regression in practice. The American Statistician, 1 (29), 3–20.

L. R. Klein and A.S. Goldberger (1964). An economic model of the United States, 1929-1952. North Holland Publishing Company, Amsterdan.

H. Theil (1971). Principles of Econometrics. John Wiley & Sons, New York.

See Also

RdetR, ki.

Examples

```
# Henri Theil's textile consumption data modified
data(theil)
head(theil)
cte = array(1,length(theil[,2]))
theil.X = cbind(cte,theil[,-(1:2)])
VIF(theil.X, TRUE, pos = 4)

# Klein and Goldberger data on consumption and wage income
data(KG)
head(KG)
cte = array(1,length(KG[,1]))
KG.X = cbind(cte,KG[,-1])
VIF(KG.X)
```