

Package ‘rvif’

October 9, 2025

Type Package

Title Collinearity Detection using Redefined Variance Inflation Factor
and Graphical Methods

Version 3.2

Date 2025-10-06

Author R. Salmerón [aut, cre],
C.B. García [aut]

Maintainer R. Salmerón <romansg@ugr.es>

Description The detection of troubling approximate collinearity in a multiple linear regression model is a classical problem in Econometrics. This package is focused on determining whether or not the degree of approximate multicollinearity in a multiple linear regression model is of concern, meaning that it affects the statistical analysis (i.e. individual significance tests) of the model. This objective is achieved by using the variance inflation factor redefined and the scatterplot between the variance inflation factor and the coefficient of variation. For more details see Salmerón R., García C.B. and García J. (2018) <[doi:10.1080/00949655.2018.1463376](https://doi.org/10.1080/00949655.2018.1463376)>, Salmerón, R., Rodríguez, A. and García C. (2020) <[doi:10.1007/s00180-019-00922-x](https://doi.org/10.1007/s00180-019-00922-x)>, Salmerón, R., García, C.B., Rodríguez, A. and García, C. (2022) <[doi:10.32614/RJ-2023-010](https://doi.org/10.32614/RJ-2023-010)>, Salmerón, R., García, C.B. and García, J. (2025) <[doi:10.1007/s10614-024-10575-8](https://doi.org/10.1007/s10614-024-10575-8)> and Salmerón, R., García, C.B., García J. (2023, working paper) <[doi:10.48550/arXiv.2005.02245](https://doi.org/10.48550/arXiv.2005.02245)>. You can also view the package vignette using 'browseVignettes(`rvif`)', the package website (<<https://www.ugr.es/local/romansg/rvif/index.html>>) using 'browseURL(system.file(`docs/index.html`, package = `rvif`))' or version control on GitHub (<https://github.com/rnoremlas/rvif_package>).

License GPL (>= 2)

Encoding UTF-8

URL <http://colldetreat.r-forge.r-project.org/>,
https://github.com/rnoremlas/rvif_package,
<https://www.ugr.es/local/romansg/rvif/index.html>

Depends R (>= 4.1.0), multiColl, car

LazyData true

Suggests knitr, rmarkdown, testthat (>= 3.0.0)
VignetteBuilder knitr
Config/testthat/edition 3
NeedsCompilation no
Repository CRAN
Date/Publication 2025-10-09 11:10:02 UTC

Contents

rvif-package	2
CDpf	3
cv_vif	4
employees	6
euribor	7
multicollinearity	8
plot.cv_vif	11
rvifs	14
SLM1	16
SLM2	16
soil	17
Wissel	18
Index	20

rvif-package	<i>Detecting multicollinearity using RVIF and graphical methods.</i>
--------------	--

Description

Detecting troubling near-multicollinearity in multiple linear regression models is a classical econometric problem. The purpose of this package is to detect it by using the Redefined Variance Inflation Factor (RVIF) and the scatterplot between the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV).

In addition, the RVIF is used to determine whether the statistical analysis of the model is affected by the degree of multicollinearity in the model.

Details

This package contains four functions. The first two, `cv_vif` and `plot.cv_vif`, respectively return the values of the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV), as well as their representation in a scatterplot. It should be noted that the VIF is useful for detecting essential multicollinearity, while the CV is useful for detecting non-essential multicollinearity. Thus, the scatterplot of both measures can provide interesting information for determining whether there is a troubling degree of multicollinearity and identifying the type of multicollinearity present and the variables causing it.

On the other hand, the function `rvif` calculates the redefined VIF and the percentage of approximate multicollinearity due to each independent variable.

Finally, multicollinearity determines whether the degree of multicollinearity in the regression model affects the statistical analysis of the model, i.e., whether the non-rejection of the null hypothesis in the individual significance tests is due to the linear relationships between the independent variables of the model.

Author(s)

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

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

References

Salmerón, R., García, C.B. and García, J. (2018). Variance inflation factor and condition number in multiple linear regression. *Journal of Statistical Computation and Simulation*, 88:2365-2384, doi: [doi:10.1080/00949655.2018.1463376](https://doi.org/10.1080/00949655.2018.1463376).

Salmerón, R., Rodríguez, A. and García, C.B. (2020). Diagnosis and quantification of the non-essential collinearity. *Computational Statistics*, 35(2), 647-666, doi: [doi:10.1007/s0018001900922-x](https://doi.org/10.1007/s0018001900922-x).

Salmerón, R., García, C.B., Rodríguez, A. and García, C. (2022). Limitations in detecting multicollinearity due to scaling issues in the `mcvis` package. *R Journal*, 14(4), 264-279, doi: [doi:10.32614/RJ2023010](https://doi.org/10.32614/RJ2023010).

Salmerón, R., García, C.B. and García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. *Computational Economics*, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Overcoming the inconsistencies of the variance inflation factor: a redefined VIF and a test to detect statistical troubling multicollinearity by Salmerón, R., García, C.B and García, J. (working paper, <https://arxiv.org/pdf/2005.02245>).

CDpf

Cobb-Douglas data

Description

Data used in Example 2 of Salmerón, García and García (2024) (subsection 4.2) on data for the Cobb-Douglas production function.

Usage

```
data("CDpf")
```

Format

A data frame containing 28 observations on the following 4 variables:

- P Production (dependent variable).
- cte Intercept.
- logK Capital (in logarithm).
- logW Work (in logarithm).

Details

This dataset was originally used by Olva Maldonado (2009).

References

Olva Maldonado, H. (2009). Análisis de la función de producción Cobb-Douglas y su aplicación en el sector productivo mexicano. Tesis, Universidad Autónoma de Chapingo.

Salmerón, R., García, C.B. and García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. Computational Economics, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Examples

```
head(CDpf, n=5)
y = CDpf[,1]
x = CDpf[,2:4]
multicollinearity(y, x)
```

cv_vif	VIF and CV calculation
--------	------------------------

Description

This function provides the values for the Variance Inflation Factor (VIF) and the Coefficient of Variation (CV) for the independent variables (excluding the intercept) in a multiple linear regression model.

Usage

```
cv_vif(x, tol = 1e-30)
```

Arguments

- x A numerical design matrix containing more than one regressor, including the intercept in the first column.
- tol A real number that indicates the tolerance beyond which the system is considered computationally unique when calculating the VIF. The default value is tol=1e-30.

Details

It is interesting to note the distinction between essential and non-essential multicollinearity. Essential multicollinearity happens when there is an approximate linear relationship between two or more independent variables (not including the intercept) while non-essential multicollinearity involves a linear relationship between the intercept and at least one independent variable. This distinction matters because the Variance Inflation Factor (VIF) only detects essential multicollinearity, while the Condition Value (CV) is useful for detecting only non-essential multicollinearity. Understanding the distinction between essential and non-essential multicollinearity and the limitations of each detection measure, can be very useful for identifying whether there is a troubling degree of multicollinearity, and determining the kind of multicollinearity present and the variables causing it.

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

- Salmerón, R., García, C.B. and García, J. (2018). Variance inflation factor and condition number in multiple linear regression. *Journal of Statistical Computation and Simulation*, 88:2365-2384, doi: [doi:10.1080/00949655.2018.1463376](https://doi.org/10.1080/00949655.2018.1463376).
- Salmerón, R., Rodríguez, A. and García, C.B. (2020). Diagnosis and quantification of the non-essential collinearity. *Computational Statistics*, 35(2), 647-666, doi: [doi:10.1007/s0018001900922-x](https://doi.org/10.1007/s0018001900922-x).
- Salmerón, R., García, C.B., Rodríguez, A. and García, C. (2022). Limitations in detecting multicollinearity due to scaling issues in the mcvis package. *R Journal*, 14(4), 264-279, doi: [doi:10.32614/RJ2023010](https://doi.org/10.32614/RJ2023010).

See Also

[plot.cv_vif](#)

Examples

```
### Example 1
### At least three independent variables, including the intercept, must be present

head(SLM1, n=5)
y = SLM1[,1]
x = SLM1[,2:3]
cv_vif(x)

### Example 2
### Creating the design matrix
```

```

library(multiColl)
set.seed(2025)
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)

### Example 3
### Obtaining the design matrix after executing the command 'lm'

library(multiColl)
set.seed(2025)
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)
u = rnorm(obs, 0, 2)
y = 5 + 4*x2 - 5*x3 + 2*x4 - x5 + u
reg = lm(y~x2+x3+x4+x5)
x = model.matrix(reg)
cv_vif(x) # identical to Example 2

### Example 3
### Computationally singular system

head(soil, n=5)
y = soil[,16]
x = soil[,-16]
cv_vif(x)

```

employees

Spanish company employee data

Description

Data used in example 3 of Salmerón, García and García (2024) (subsection 4.3) on the number of employees of Spanish companies.

Usage

```
data("employees")
```

Format

A data frame with 15 observations on the following 5 variables:

NE Number of employees (dependent variable).

cte Intercept.

FA Fixed assets (in euros).

OI Operating income (in euros).

S Sales (in euros).

Details

This dataset is originally used by Salmerón, Rodríguez, García and García (2020).

References

Salmerón, R., Rodríguez, A., García, C.B. and García, J. (2020). The VIF and MSE in raise regression. *Mathematics*, 8(4), doi: [doi:10.3390/math8040605](https://doi.org/10.3390/math8040605).

Salmerón, R., García, C.B. and García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. *Computational Economics*, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Examples

```
head(employees, n=5)
y = employees[,1]
x = employees[,3:5]
multicollinearity(y, x)
```

euribor

Euribor data

Description

Data used in example 1 of Salmerón, García and García (2024) (subsection 4.1) on Euribor data.

Usage

```
data("euribor")
```

Format

A data frame with 47 observations on the following 5 variables:

E Euribor (dependent variable, in percentage).

cte Intercept.

HIPC Harmonized index of consumer prices (in percentage).

BC Balance of payments to net current account (millions of euros).

GD Government deficit to net nonfinancial accounts (millions of euros).

Details

This dataset is originally used by Salmerón, Rodríguez and García (2020).

References

Salmerón, R., Rodríguez, A. and García, C.B. (2020). Diagnosis and quantification of the non-essential collinearity. Computational Statistics, 35(2), 647-666, doi: [doi:10.1007/s0018001900922-x](https://doi.org/10.1007/s0018001900922-x).

Salmerón, R., García, C.B. and García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. Computational Economics, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Examples

```
head(euribor, n=5)
y = euribor[,1]
x = euribor[,2:5]
multicollinearity(y, x)
```

multicollinearity	<i>Decision Rule to Detect Troubling Multicollinearity</i>
-------------------	--

Description

Given a multiple linear regression model with n observations and k independent variables, the degree of near-multicollinearity affects its statistical analysis (with a level of significance of alpha%) if there is a variable i, with i = 1,...,k, that verifies that the null hypothesis is not rejected in the original model and is rejected in the orthogonal model of reference.

Usage

```
multicollinearity(y, x, alpha = 0.05)
```

Arguments

- y A numerical vector representing the dependent variable of the model.
- x A numerical design matrix that should contain more than one regressor (intercept included in the first column).
- alpha Significance level (by default, 5%).

Details

This function compares the individual inference of the original model with that of the orthonormal model taken as reference.

Thus, if the null hypothesis is rejected in the individual significance tests in the model where there are no linear relationships between the independent variables (orthonormal) and is not rejected in the original model, the reason for the non-rejection is due to the existing linear relationships between the independent variables (multicollinearity) in the original model.

The second model is obtained from the first model by performing a QR decomposition, which eliminates the initial linear relationships.

Value

The function returns the value of the RVIF and the established thresholds, as well as indicating whether or not the individual significance analysis is affected by multicollinearity at the chosen significance level.

Author(s)

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

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

References

Salmerón, R., García, C.B. and García, J. (2025). A Redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. Computational Economics, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Overcoming the inconsistencies of the variance inflation factor: a redefined VIF and a test to detect statistical troubling multicollinearity by Salmerón, R., García, C.B and García, J. (working paper, <https://arxiv.org/pdf/2005.02245>).

See Also

[rvifs](#)

Examples

```
### Example 1

set.seed(2024)
obs = 100
cte = rep(1, obs)
x2 = rnorm(obs, 5, 0.01) # related to intercept: non essential
x3 = rnorm(obs, 5, 10)
x4 = x3 + rnorm(obs, 5, 0.5) # related to x3: essential
x5 = rnorm(obs, -1, 3)
x6 = rnorm(obs, 15, 0.5)
y = 4 + 5*x2 - 9*x3 - 2*x4 + 2*x5 + 7*x6 + rnorm(obs, 0, 2)
x = cbind(cte, x2, x3, x4, x5, x6)
```

```

    multicollinearity(y, x)

### Example 2
### Effect of sample size

    obs = 25 # by decreasing the number of observations affected to x4
    cte = rep(1, obs)
    x2 = rnorm(obs, 5, 0.01) # related to intercept: non essential
    x3 = rnorm(obs, 5, 10)
    x4 = x3 + rnorm(obs, 5, 0.5) # related to x3: essential
    x5 = rnorm(obs, -1, 3)
    x6 = rnorm(obs, 15, 0.5)
    y = 4 + 5*x2 - 9*x3 - 2*x4 + 2*x5 + 7*x6 + rnorm(obs, 0, 2)
    x = cbind(cte, x2, x3, x4, x5, x6)
    multicollinearity(y, x)

### Example 3

    y = 4 - 9*x3 - 2*x5 + rnorm(obs, 0, 2)
    x = cbind(cte, x3, x5) # independently generated
    multicollinearity(y, x)

### Example 4
### Detection of multicollinearity in Wissel data

    head(Wissel, n=5)
    y = Wissel[,2]
    x = Wissel[,3:6]
    multicollinearity(y, x)

### Example 5
### Detection of multicollinearity in euribor data

    head(euribor, n=5)
    y = euribor[,1]
    x = euribor[,2:5]
    multicollinearity(y, x)

### Example 6
### Detection of multicollinearity in Cobb-Douglas production function data

    head(CDpf, n=5)
    y = CDpf[,1]
    x = CDpf[,2:4]
    multicollinearity(y, x)

### Example 7
### Detection of multicollinearity in number of employees of Spanish companies data

    head(employees, n=5)
    y = employees[,1]
    x = employees[,3:5]
    multicollinearity(y, x)

```

```

### Example 8
### Detection of multicollinearity in simple linear model simulated data

head(SLM1, n=5)
y = SLM1[,1]
x = SLM1[,2:3]
multicollinearity(y, x)

head(SLM2, n=5)
y = SLM2[,1]
x = SLM2[,2:3]
multicollinearity(y, x)

### Example 9
### Detection of multicollinearity in soil characteristics data

head(soil, n=5)
y = soil[,16]
x = soil[,,-16]
x = cbind(rep(1, length(y)), x) # the design matrix has to have the intercept in the first column
multicollinearity(y, x)
multicollinearity(y, x[,,-3]) # eliminating the problematic variable (SumCation)

### Example 10
### The intercept must be in the first column of the design matrix

set.seed(2025)
obs = 100
cte = rep(1, obs)
x2 = sample(1:500, obs)
x3 = sample(1:500, obs)
x4 = rep(4, obs)
x = cbind(cte, x2, x3, x4)
u = rnorm(obs, 0, 2)
y = 5 + 2*x2 - 3*x3 + 10*x4 + u
multicollinearity(y, x)
multicollinearity(y, x[,,-4]) # the constant variable is removed

```

plot.cv_vif

Scatterplot of CV vs VIF

Description

This function provides a graphical representation of a scatter plot showing the Coefficient of Variation (CV) and the Variance Inflation Factor (VIF) for the independent variables (excluding the intercept) of a multiple linear regression model.

Usage

```
## S3 method for class 'cv_vif'
plot(x, ...)
```

Arguments

`x` This is the output of the function `cv_vif`.
`...` other graphical parameters.

Details

The distinction between essential and non-essential multicollinearity and the limitations of each measure (CV and VIF) for detecting the different kinds of multicollinearity, can be very useful for identifying whether there is a troubling degree of multicollinearity, and determining the kind of multicollinearity present and the variables causing it.

For this purpose, it is important to include the lines corresponding to the established thresholds for each measure in the representation of the scatter plot of the CV and VIF: a dashed vertical line for 0.1002506 (CV) and a dotted horizontal line for 10 (VIF). These lines determine four regions (see Example 1), which 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).

Author(s)

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

References

Salmerón, R., García, C.B. and García, J. (2018). Variance inflation factor and condition number in multiple linear regression. *Journal of Statistical Computation and Simulation*, 88:2365-2384, doi: [doi:10.1080/00949655.2018.1463376](https://doi.org/10.1080/00949655.2018.1463376).

Salmerón, R., Rodríguez, A. and García, C.B. (2020). Diagnosis and quantification of the non-essential collinearity. *Computational Statistics*, 35(2), 647-666, doi: [doi:10.1007/s0018001900922-x](https://doi.org/10.1007/s0018001900922-x).

Salmerón, R., García, C.B., Rodríguez, A. and García, C. (2022). Limitations in detecting multicollinearity due to scaling issues in the *mcvis* package. *R Journal*, 14(4), 264-279, doi: [doi:10.32614/RJ2023010](https://doi.org/10.32614/RJ2023010).

See Also

[cv_vif](#)

Examples

```
### Example 1

plot(-2:20, -2:20, type = "n", xlab="Coefficient of Variation",
```

```

                                ylab="Variance Inflation Factor")
abline(h=10, col="red", lwd=3, lty=2)
abline(h=0, col="black", lwd=1)
abline(v=0.1002506, col="red", lwd=3, lty=3)
#abline(v=0, col="red", lwd=1)
text(-1.25, 2, "A", pos=3, col="blue")
text(-1.25, 12, "B", pos=3, col="blue")
text(10, 12, "C", pos=3, col="blue")
text(10, 2, "D", pos=3, col="blue")

### Example 2

library(multiColl)
set.seed(2025)
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)

plot(cv_vif(x))
abline(h=10, col="red", lwd=2, lty=2)
abline(h=0, col="black", lwd=1)
abline(v=0.1002506, col="red", lwd=2, lty=3)
  labels = c()
  for(i in 1:length(cv_vif(x)[[1]])){labels = c(labels, i+1)}
  text(cv_vif(x)[[1]], cv_vif(x)[[2]], labels = labels, pos=3)

cv_vif(x) |> plot()
abline(h=10, col="red", lwd=2, lty=2)
abline(h=0, col="black", lwd=1)
abline(v=0.1002506, col="red", lwd=2, lty=3)
  labels = c()
  for(i in 1:length(cv_vif(x)[[1]])){labels = c(labels, i+1)}
  text(cv_vif(x)[[1]], cv_vif(x)[[2]], labels = labels, pos=3)

### Example 3
### Graphical representation is not possible

head(SLM2, n=5)
x = SLM2[,2:3]
plot(cv_vif(x))
cv_vif(x) |> plot()

### Example 4
### Computationally singular system

head(soil, n=5)
x = soil[, -16]
plot(cv_vif(x))
cv_vif(x) |> plot()

```

rvifs	<i>RVIF calculation</i>
-------	-------------------------

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

```
rvifs(x, ul = TRUE, intercept = TRUE, tol = 1e-30)
```

Arguments

x	A numerical design matrix that should contain more than one regressor. If it has an intercept, this must be in the first column of the matrix).
ul	A logical value that indicates if the variables in the design matrix x are transformed to unit length. By default ul=TRUE.
intercept	A logical value that indicates if the design matrix x has an intercept. By default intercept=TRUE.
tol	Value determining whether the system is computationally singular. By default tol=1e-30.

Details

The Redefined Variation Inflation Factor (RVIF) is capable to detect both kind of multicollinearity: the essential (approximate linear relationship between at least two independent variables excluding the intercept) and non-essential (approximate 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.

Author(s)

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

References

R. Salmerón, C. García, and J. García. (2018). Variance inflation factor and condition number in multiple linear regression. Journal of Statistical Computation and Simulation, 88:2365-2384, doi: [doi:10.1080/00949655.2018.1463376](https://doi.org/10.1080/00949655.2018.1463376).

Salmerón, R., Rodríguez, A. and García, C.B. (2020). Diagnosis and quantification of the non-essential collinearity. *Computational Statistics*, 35(2), 647-666, doi: [doi:10.1007/s0018001900922-x](https://doi.org/10.1007/s0018001900922-x).

Salmerón, R., García, C.B. y García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. *Computational Economics*, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Examples

Example 1

```
library(multiColl)
set.seed(2025)
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)
rvifs(x)
```

Example 2

The special case of the simple linear regression model

```
head(SLM1, n=5)
x = SLM1[,2:3]
rvifs(x)
```

Example 3

The intercept must be in the first column of the design matrix

```
set.seed(2025)
obs = 100
cte = rep(1, obs)
x2 = sample(1:500, obs)
x3 = sample(1:500, obs)
x4 = rep(4, obs)
x = cbind(cte, x2, x3, x4)
rvifs(x) # also: perfect multicollinearity between the intercept and the constant variable
rvifs(x[,-1], intercept = FALSE) # removing the constant from the design matrix
```

Example 4

Cases of perfect multicollinearity or computationally singular systems

```
head(soil, n=5)
x = soil[,-16]
rvifs(x)
```

SLM1

*First simulated data for the simple linear regression model***Description**

First data used in example 4 of Salmerón, García and García (2024) (subsection 4.4) on the special case of the simple linear model.

Usage

```
data("SLM1")
```

Format

A data frame with 50 observations on the following 3 variables:

y1 Dependent variable simulated as $y = 3 + 4 \cdot V + u$ where u is normally distributed with a mean of 0 and a variance of 2.

cte Intercept.

V Simulated from a normal distribution with a mean of 10 and a variance of 100.

References

Salmerón, R., García, C.B. and García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. Computational Economics, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Examples

```
head(SLM1, n=5)
y = SLM1[,1]
x = SLM1[,2:3]
multicollinearity(y, x)
```

SLM2

*Second simulated data for the simple linear regression model***Description**

Second data used in example 4 of Salmerón, García and García (2024) (subsection 4.4) on the special case of the simple linear model.

Usage

```
data("SLM2")
```


Format

A data frame with 50 observations on the following 3 variables:

y2 Dependent variable simulated as $y = 3 + 4 \cdot Z + u$ where u is normally distributed with a mean of 0 and a variance of 2.

cte Intercept.

Z Simulated from a normal distribution with a mean of 10 and a variance of 0.1.

References

Salmerón, R., García, C.B. and García, J. (2025). A redefined Variance Inflation Factor: overcoming the limitations of the Variance Inflation Factor. *Computational Economics*, 65, 337-363, doi: [doi:10.1007/s10614024105758](https://doi.org/10.1007/s10614024105758).

Examples

```
head(SLM2, n=5)
y = SLM2[,1]
x = SLM2[,2:3]
multicollinearity(y, x)
```

soil	<i>Soil characteristics data</i>
------	----------------------------------

Description

Data used in Bondell and Reich's paper on soil characteristics used as predictors of forest diversity.

Usage

```
data("soil")
```

Format

A data frame with 20 observations on the following 16 variables.

BaseSat % Base Saturation.

SumCation Sum Cations (sums of cations like calcium, magnesium, potassium and sodium).

CECbuffer CEC.

Ca Calcium.

Mg Magnesium.

K Potassium.

Na Sodium.

P Phosphorus.

Cu Copper.

Zn Zinc.
Mn Manganese.
HumicMatter Humic Matter.
Density Density.
pH pH.
ExchAc Exchangeable Acidity.
Diversity Forest diversity (dependent variable).

Details

This dataset is originally used by Bondell and Reich (2008).

References

Bondell, H.D. and Reich. B.J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. Biometrics, 64 (1), 115–23, doi: [doi:10.1111/j.15410420.2007.00843.x](https://doi.org/10.1111/j.15410420.2007.00843.x).

Examples

```
head(soil, n=5)
y = soil[,16]
x = soil[,-16]
x = cbind(rep(1, length(y)), x) # the design matrix has to have the intercept in the first column
multicollinearity(y, x)
multicollinearity(y, x[,-3]) # eliminating the problematic variable (SumCation)
```

Wissel	Wissel data
--------	-------------

Description

Wissel data on outstanding mortgage debt.

Usage

```
data("Wissel")
```

Format

A data frame with 17 observations on the following 6 variables:

- t Year.
- D Outstanding mortgage debt (dependent variable).
- cte Intercept.
- C Personal consumption (trillions of dollars).
- I Personal income (trillions of dollars).
- CP Outstanding consumer credit (trillions of dollars).

References

Wissel, J. (2009). A new biased estimator for multivariate regression models with highly collinear variables. Ph.D. thesis, Erlangung des naturwissenschaftlichen Doktorgrades der Bayerischen Julius-Maximilians-Universität Würzburg, url: <https://opus.bibliothek.uni-wuerzburg.de/opus4-wuerzburg/frontdoor/deliver/index/docId/2949/file/wissel.pdf>.

Examples

```
head(Wissel, n=5)
y = Wissel[,2]
x = Wissel[,3:6]
multicollinearity(y, x)
```

Index

- * **Coefficient of Variation**
 - cv_vif, 4
 - plot.cv_vif, 11
 - rvif-package, 2
 - rvifs, 14
 - * **Individual significance tests**
 - multicollinearity, 8
 - * **Orthonormal model**
 - multicollinearity, 8
 - * **Redefined Variation Inflation Factor**
 - rvif-package, 2
 - * **Scatter plot**
 - plot.cv_vif, 11
 - * **Variance inflation factor redefined**
 - multicollinearity, 8
 - * **Variance inflation factor**
 - cv_vif, 4
 - plot.cv_vif, 11
 - * **Variation Inflation Factor Redefined**
 - rvifs, 14
 - * **Variation Inflation Factor**
 - rvif-package, 2
 - rvifs, 14
 - * **datasets**
 - CDpf, 3
 - employees, 6
 - euribor, 7
 - SLM1, 16
 - SLM2, 16
 - soil, 17
 - Wissel, 18
- CDpf, 3
- cv_vif, 4, 12
- employees, 6
- euribor, 7
- multicollinearity, 8
- plot.cv_vif, 5, 11
- rvif (rvif-package), 2
- rvif-package, 2
- rvifs, 9, 14
- SLM1, 16
- SLM2, 16
- soil, 17
- Wissel, 18