# Package ‘india’

May 4, 2025

**Type** Package

**Title** Influence Diagnostics in Statistical Models

**Version** 0.1-1

**Date** 2025-05-03

**Maintainer** Felipe Osorio <faosorios.stat@gmail.com>

**Description** Set of routines for influence diagnostics by using case-deletion in ordinary least squares, nonlinear regression [Ross (1987). <doi:10.2307/3315198>], ridge estimation [Walker and Birch (1988). <doi:10.1080/00401706.1988.10488370>] and least absolute deviations (LAD) regression [Sun and Wei (2004). <doi:10.1016/j.spl.2003.08.018>].

**Depends** R(>= 3.5.0), fastmatrix, L1pack

**Imports** stats



**URL** https://github.com/faosorios/india

**NeedsCompilation** yes

**LazyLoad** yes

**Author** Felipe Osorio [aut, cre] (ORCID: <https://orcid.org/0000-0002-4675-5201>)

**Repository** CRAN

**Date/Publication** 2025-05-03 22:30:02 UTC

## Contents

cooks.distance *Cook's distances*

## Description

Cook's distance is a measure to assess the influence of the *i*th observation on the model parameter estimates. This function computes the Cook's distance based on leave-one-out cases deletion for ordinary least squares, nonlinear least squares, lad and ridge regression.

## Usage

```
## S3 method for class 'lad'
cooks.distance(model, ...)
## S3 method for class 'nls'
cooks.distance(model, ...)
## S3 method for class 'ols'
cooks.distance(model, ...)
## S3 method for class 'ridge'
cooks.distance(model, type = "cov", ...)
```

## Arguments

| | |
|---|---|
| `model` | an R object, returned by `ols`, `nls`, `lad` or `ridge`. |
| `type` | only required for `'ridge'` objects, options available are `"1st"`, `"cov"` and `"both"` to obtain the Cook's distance based on Equation (2.5), (2.6) or both by Walker and Birch (1988), respectively. |
| `...` | further arguments passed to or from other methods. |

## Value

A vector whose *i*th element contains the Cook's distance,

$$D_i(\boldsymbol{M}, c) = \frac{(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^T \boldsymbol{M} (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})}{c},$$

for $i = 1, \dots, n$, with $\boldsymbol{M}$ a positive definite matrix and $c > 0$. Specific choices of $\boldsymbol{M}$ and $c$ are done for objects of class `ols`, `nls`, `lad` and `ridge`.

The Cook's distance for nonlinear regression is based on linear approximation, which may be inappropriate for expectation surfaces markedly nonplanar.

## References

Cook, R.D., Weisberg, S. (1980). Characterizations of an empirical influence function for detecting influential cases in regression. *Technometrics* **22**, 495-508. doi:10.1080/00401706.1980.10486199

Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.

Ross, W.H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression. *The Canadian Journal of Statistics* **15**, 91-103. doi:10.2307/3315198

Sun, R.B., Wei, B.C. (2004). On influence assessment for LAD regression. *Statistics & Probability Letters* **67**, 97-110. doi:10.1016/j.spl.2003.08.018

Walker, E., Birch, J.B. (1988). Influence measures in ridge regression. *Technometrics* **30**, 221-227. doi:10.1080/00401706.1988.10488370

## Examples

```
# Cook's distances for linear regression
fm <- ols(stack.loss ~ ., data = stackloss)
CD <- cooks.distance(fm)
plot(CD, ylab = "Cook's distances", ylim = c(0,0.8))
text(21, CD[21], label = as.character(21), pos = 3)

# Cook's distances for LAD regression
fm <- lad(stack.loss ~ ., data = stackloss)
CD <- cooks.distance(fm)
plot(CD, ylab = "Cook's distances", ylim = c(0,0.4))
text(17, CD[17], label = as.character(17), pos = 3)

# Cook's distances for ridge regression
data(portland)
fm <- ridge(y ~ ., data = portland)
CD <- cooks.distance(fm)
plot(CD, ylab = "Cook's distances", ylim = c(0,0.5))
text(8, CD[8], label = as.character(8), pos = 3)

# Cook's distances for nonlinear regression
data(skeena)
model <- recruits ~ b1 * spawners * exp(-b2 * spawners)
fm <- nls(model, data = skeena, start = list(b1 = 3, b2 = 0))
CD <- cooks.distance(fm)
plot(CD, ylab = "Cook's distances", ylim = c(0,0.35))
obs <- c(5, 6, 9, 19, 25)
text(obs, CD[obs], label = as.character(obs), pos = 3)
```

---

`leverages` *Leverages*

---

## Description

Computes leverage measures from a fitted model object.

## Usage

```
leverages(model, ...)
## S3 method for class 'lm'
leverages(model, infl = lm.influence(model, do.coef = FALSE), ...)
```

```
## S3 method for class 'nls'
leverages(model, ...)
## S3 method for class 'ols'
leverages(model, ...)
## S3 method for class 'ridge'
leverages(model, ...)

## S3 method for class 'nls'
hatvalues(model, ...)
## S3 method for class 'ols'
hatvalues(model, ...)
## S3 method for class 'ridge'
hatvalues(model, ...)
```

## Arguments

| | |
|---|---|
| `model` | an R object, returned by `lm`, `nls`, `ols` or `ridge`. |
| `infl` | influence structure as returned by `lm.influence`. |
| `...` | further arguments passed to or from other methods. |

## Value

A vector containing the diagonal of the prediction (or 'hat') matrix.

For linear regression (i.e., for `"lm"` or `"ols"` objects) the prediction matrix assumes the form

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T,$$

in which case, $h_{ii} = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x}_i$ for $i = 1, \dots, n$. Whereas for ridge regression, the prediction matrix is given by

$$\boldsymbol{H}(\lambda) = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T,$$

where $\lambda$ represents the ridge parameter. Thus, the diagonal elements of $\boldsymbol{H}(\lambda)$, are $h_{ii}(\lambda) = \boldsymbol{x}_i^T(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{x}_i$, $i = 1, \dots, n$.

In nonlinear regression, the tangent plane leverage matrix is given by

$$\hat{\boldsymbol{H}} = \hat{\boldsymbol{F}}(\hat{\boldsymbol{F}}^T\hat{\boldsymbol{F}})^{-1}\hat{\boldsymbol{F}}^T,$$

where $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{\beta})$ is the $n \times p$ local model matrix with $i$th row $\partial f_i(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ and $\hat{\boldsymbol{F}} = \boldsymbol{F}(\hat{\boldsymbol{\beta}})$.

## Note

This function never creates the prediction matrix and only obtains its diagonal elements from the singular value decomposition of $\boldsymbol{X}$ or $\hat{\boldsymbol{F}}$.

Function `hatvalues` only is a wrapper for function `leverages`.

## References


Chatterjee, S., Hadi, A.S. (1988). *Sensivity Analysis in Linear Regression*. Wiley, New York.

Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.

Ross, W.H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression. *The Canadian Journal of Statistics* **15**, 91-103. doi:10.2307/3315198

St. Laurent, R.T., Cook, R.D. (1992). Leverage and superleverage in nonlinear regression. *Journal of the Amercian Statistical Association* **87**, 985-990. doi:10.1080/01621459.1992.10476253

Walker, E., Birch, J.B. (1988). Influence measures in ridge regression. *Technometrics* **30**, 221-227. doi:10.1080/00401706.1988.10488370


## Examples

```
# Leverages for linear regression
fm <- ols(stack.loss ~ ., data = stackloss)
lev <- leverages(fm)
cutoff <- 2 * mean(lev)
plot(lev, ylab = "Leverages", ylim = c(0,0.45))
abline(h = cutoff, lty = 2, lwd = 2, col = "red")
text(17, lev[17], label = as.character(17), pos = 3)

# Leverages for ridge regression
data(portland)
fm <- ridge(y ~ ., data = portland)
lev <- leverages(fm)
cutoff <- 2 * mean(lev)
plot(lev, ylab = "Leverages", ylim = c(0,0.7))
abline(h = cutoff, lty = 2, lwd = 2, col = "red")
text(10, lev[10], label = as.character(10), pos = 3)

# Leverages for nonlinear regression
data(skeena)
model <- recruits ~ b1 * spawners * exp(-b2 * spawners)
fm <- nls(model, data = skeena, start = list(b1 = 3, b2 = 0))
lev <- leverages(fm)
plot(lev, ylab = "Leverages", ylim = c(0,0.25))
obs <- c(1,9)
text(obs, lev[obs], label = as.character(obs), pos = 3)
```

---

# `logLik.displacement` *Likelihood Displacement*

---

## Description

Compute the likelihood displacement influence measure based on leave-one-out cases deletion for linear models, lad and ridge regression.

## Usage

```
logLik.displacement(model, ...)
## S3 method for class 'lm'
logLik.displacement(model, pars = "full", ...)
## S3 method for class 'nls'
logLik.displacement(model, ...)
## S3 method for class 'ols'
logLik.displacement(model, pars = "full", ...)
## S3 method for class 'lad'
logLik.displacement(model, method = "quasi", pars = "full", ...)
## S3 method for class 'ridge'
logLik.displacement(model, pars = "full", ...)
```

## Arguments

| | |
|---|---|
| `model` | an R object, returned by `lm`, `nls`, `ols`, `lad` or `ridge`. |
| `pars` | should be considered the whole vector of parameters (`pars = "full"`), or only the vector of coefficients (`pars = "coef"`). This option is not used for `nls` objects. |
| `method` | only required for `'lad'` objects, options available are `"quasi"` and `"BR"` to obtain the likelihood displacement based on Sun and Wei (2004) and Elian et al. (2000) approaches, respectively. |
| `...` | further arguments passed to or from other methods. |

## Value

A vector whose *i*th element contains the distance between the likelihood functions,

$$LD_i(\boldsymbol{\beta}, \sigma^2) = 2\{l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) - l(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2_{(i)})\},$$

for `pars = "full"`, where $\hat{\boldsymbol{\beta}}_{(i)}$ and $\hat{\sigma}^2_{(i)}$ denote the estimates of $\boldsymbol{\beta}$ and $\sigma^2$ when the *i*th observation is removed from the dataset. If we are interested only in $\boldsymbol{\beta}$ (i.e. `pars = "coef"`) the likelihood displacement becomes

$$LD_i(\boldsymbol{\beta}|\sigma^2) = 2\{l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) - \max_{\sigma^2} l(\hat{\boldsymbol{\beta}}_{(i)}, \hat{\sigma}^2)\}.$$

## References

Cook, R.D., Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman and Hall, London.

Cook, R.D., Pena, D., Weisberg, S. (1988). The likelihood displacement: A unifying principle for influence measures. *Communications in Statistics - Theory and Methods* **17**, 623-640. doi:10.1080/03610928808829645

Elian, S.N., Andre, C.D.S., Narula, S.C. (2000). Influence measure for the L1 regression. *Communications in Statistics - Theory and Methods* **29**, 837-849. doi:10.1080/03610920008832518

Ogueda, A., Osorio, F. (2025). Influence diagnostics for ridge regression using the Kullback-Leibler divergence. *Statistical Papers* **66**, 85. doi:10.1007/s00362025017011

Ross, W.H. (1987). The geometry of case deletion and the assessment of influence in nonlinear regression. *The Canadian Journal of Statistics* **15**, 91-103. doi:10.2307/3315198

Sun, R.B., Wei, B.C. (2004). On influence assessment for LAD regression. *Statistics & Probability Letters* **67**, 97-110. doi:10.1016/j.spl.2003.08.018

## Examples

```
# Likelihood displacement for linear regression
fm <- ols(stack.loss ~ ., data = stackloss)
LD <- logLik.displacement(fm)
plot(LD, ylab = "Likelihood displacement", ylim = c(0,9))
text(21, LD[21], label = as.character(21), pos = 3)

# Likelihood displacement for LAD regression
fm <- lad(stack.loss ~ ., data = stackloss)
LD <- logLik.displacement(fm)
plot(LD, ylab = "Likelihood displacement", ylim = c(0,1.5))
text(17, LD[17], label = as.character(17), pos = 3)

# Likelihood displacement for ridge regression
data(portland)
fm <- ridge(y ~ ., data = portland)
LD <- logLik.displacement(fm)
plot(LD, ylab = "Likelihood displacement", ylim = c(0,4))
text(8, LD[8], label = as.character(8), pos = 3)

# Likelihood displacement for nonlinear regression
data(skeena)
model <- recruits ~ b1 * spawners * exp(-b2 * spawners)
fm <- nls(model, data = skeena, start = list(b1 = 3, b2 = 0))
LD <- logLik.displacement(fm)
plot(LD, ylab = "Likelihood displacement", ylim = c(0,0.7))
obs <- c(5, 6, 9, 19, 25)
text(obs, LD[obs], label = as.character(obs), pos = 3)
```

---

# portland *Portland cement dataset*

---

## Description

This dataset comes from an experimental investigation of the heat evolved during the setting and hardening of Portland cements of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was produced.

## Usage

```
data(portland)
```

## Format

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

**y** The heat evolved after 180 days of curing, measured in calories per gram of cement.

**x1** Tricalcium aluminate.

**x2** Tricalcium silicate.

**x3** Tetracalcium aluminoferrite.

**x4** $\beta$-dicalcium silicate.

## Source

Kaciranlar, S., Sakallioglu, S., Akdeniz, F., Styan, G.P.H., Werner, H.J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on Portland cement. *Sankhya, Series B* **61**, 443-459.

---

# `relative.condition` *Relative change in the condition number*

---

## Description

Compute the relative condition index to identify collinearity-influential points in linear models.

## Usage

```
relative.condition(x)
```

## Arguments

| | |
|---|---|
| `x` | the model matrix $\boldsymbol{X}$. |

## Value

To assess the influence of the *i*th row of $\boldsymbol{X}$ on the condition index of $\boldsymbol{X}$, Hadi (1988) proposed the relative change,

$$\delta_i = \frac{\kappa_{(i)} - \kappa}{\kappa},$$

for $i = 1, \dots, n$, where $\kappa = \kappa(\boldsymbol{X})$ and $\kappa_{(i)} = \kappa(\boldsymbol{X}_{(i)})$ denote the (scaled) condition index for $\boldsymbol{X}$ and $\boldsymbol{X}_{(i)}$, respectively.

## References

Chatterjee, S., Hadi, A.S. (1988). *Sensivity Analysis in Linear Regression*. Wiley, New York.

Hadi, A.S. (1988). Diagnosing collinerity-influential observations. *Computational Statistics & Data Analysis* **7**, 143-159. doi:10.1016/01679473(88)900898.

## Examples

```
data(portland)
fm <- ridge(y ~ ., data = portland, x = TRUE)
x <- fm$x
rel <- relative.condition(x)
plot(rel, ylab = "Relative condition number", ylim = c(-0.1,0.4))
abline(h = 0, lty = 2, lwd = 2, col = "red")
text(3, rel[3], label = as.character(3), pos = 3)
```

---

skeena *Skeena River sockeye salmon data*

---

## Description

The data have 28 observations of spawners and recruits (units are thousands of fish) from 1940 until 1967 for the Skeena river sockeye salmon stock.

## Usage

```
data(skeena)
```

## Format

A data frame with 28 observations on the following 3 variables.

**year** Years in which the number of spawners and recruits were recorded.

**spawners** Size of the annual spawning stock.

**recruits** Production of new catchable-sized fish.

## Source

Carroll, R.J., Ruppert, D. (1988). *Transformation and Weighting in Regression*. Chapman and Hall, London.

# Index