

# Package ‘Routliers’

April 28, 2019

**Title** Robust Outliers Detection

**Version** 0.0.0.1

**Description** Detecting outliers using robust methods,  
i.e. the Median Absolute Deviation (MAD) for univariate out-  
liers; Leys, Ley, Klein, Bernard, & Licata (2013) <doi:10.1016/j.jesp.2013.03.013>  
and the Mahalanobis-Minimum Covariance Determinant (MMCD) for multivariate out-  
liers; Leys, C., Klein, O., Dominicy, Y. & Ley, C. (2018) <doi:10.1016/j.jesp.2017.09.011>.  
There is also the more known but less robust Mahalanobis distance method, only for compari-  
son purposes.

**License** MIT + file LICENSE

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 6.1.1

**Depends** R (>= 2.10)

**BugReports** <https://github.com/mdelacore/Routliers/issues>

**Suggests** knitr, rmarkdown, testthat

**Imports** MASS, stats, graphics

**NeedsCompilation** no

**Author** Olivier Klein [aut],  
Marie Delacre [aut, cre]

**Maintainer** Marie Delacre <marie.delacre@ulb.ac.be>

**Repository** CRAN

**Date/Publication** 2019-04-28 08:50:02 UTC

## R topics documented:

Attacks . . . . .	2
Intention . . . . .	4
Morality . . . . .	5
outliers_mad . . . . .	6

outliers_mahalanobis . . . . .	7
outliers_mcd . . . . .	8
plot_outliers_mad . . . . .	9
plot_outliers_mahalanobis . . . . .	9
plot_outliers_mcd . . . . .	10

<b>Index</b>	<b>12</b>
--------------	-----------

---

Attacks	<i>Data collected the day after the terrorist attacks in Brussels (on the morning of 22 March 2016) assessing the Sense of Coherence, anxiety and depression symptoms of 2077 subjects (1056 were in Brussels during the terrorist attacks, and 1021 were not).</i>
---------	---

---

### Description

The Sense of Coherence was assessed with the SOC-13 (Antonovsky, 1987): 7-point Likert scale (13 items) Anxiety and depression were assessed with the HSCL-25 (Derogatis, Lipman, Rickels, Uhlenhuth & Covi, 1974). Subjects have to mention in a 4-point Likert Scale how much there were bothered or upset by each trouble during the last 14 days (1 = not at all; 2 = a little; quite a few; 4 = a lot).

### Usage

```
data(Attacks)
```

### Format

A data frame with 2077 rows and 46 variables:

**age** age of participants, in years

**presencebxl** were participants present in Brussels during the terrorist attacks; 1 = yes; -1 = no

**genre** participant gender, 1 = female; -1 = male

**soc1** Vous avez le sentiment que vous ne vous souciez pas reellement de ce qui se passe autour de vous: 1 = Tres rarement ou rarement; 7 = Souvent

**soc1r** item1 reversed

**soc2** Vous est-il arrive dans le passe d etre surpris(e) par le comportement de gens que vous pensiez connaitre tres bien ?: 1 = Jamais; 7 = Toujours

**soc2r** item2 reversed

**soc3** Est-il arrive que des gens sur lesquels vous comptiez vous decoivent ?: 1= Jamais; 7 = Toujours

**soc3r** sense of coherence, item3 reversed

**soc4** Jusqu a maintenant, votre vie : 1 = N a eu aucun but ni objectif clair; 7 = A eu des buts et des objectifs tres clairs

**soc5** Avez-vous le sentiment que vous etes traite(e) injustement ?:1 = Tres souvent; 7 = Tres rarement ou jamais

- soc6** Avez-vous le sentiment que vous etes dans une situation inconnue et que vous ne savez pas quoi faire ? : 1 = Tres souvent; 7 = Tres rarement ou jamais
- soc7** Faire les choses que vous faites quotidiennement est : 1 = Une source de plaisir et de satisfaction; 7 = Une source de souffrance profonde et d ennui
- soc7r** item7 reversed
- soc8** Avez-vous des idees ou des sentiments confus(es) ? : 1 = Tres souvent; 7 = Tres rarement ou jamais
- soc9** Vous arrive-t-il d avoir des sentiments intimes que vous prefereriez ne pas avoir ? : 1 = Tres souvent; 7 = Tres rarement ou jamais
- soc10** Beaucoup de gens (meme s'ils ont beaucoup de caractere) se sentent parfois de pauvres cloches. Avez-vous deja eu ce sentiment dans le passe ? : 1 = Jamais; 7 = Tres souvent
- soc10r** item10 reversed
- soc11** Quand quelque chose arrive, vous trouvez generalement que : 1 = Vous surestimez ou sous-estimez son importance; 7 = Vous voyez les choses dans de justes proportions
- soc12** Avez-vous le sentiment que les choses que vous faites dans la vie quotidienne ont peu de sens ? : 1 = Tres souvent; 7 = Tres rarement ou jamais
- soc13** Vous avez le sentiment que vous n etes pas sur(e) de vous maitriser : 1 = Tres souvent; 7 = Tres rarement ou jamais
- hsc1** Mal de tete
- hsc2** Tremblement
- hsc3** Fatigue ou etourdissement
- hsc4** Nervosite, agitation au fond de soi
- hsc5** Peur soudaine sans raison particuliere
- hsc6** Continuellement peureux ou anxieux
- hsc7** Battements du coeur qui s'emballent
- hsc8** Sensation d etre tendu, stresse
- hsc9** Crise d angoisse ou de panique
- hsc10** Tellement agite qu'il en est difficile de rester assis
- hsc11** Manque d energie, tout va plus lentement que d habitude
- hsc12** Se fait facilement des reproches
- hsc13** Pleure facilement
- hsc14** Pense a se tuer
- hsc15** Mauvais appetit
- hsc16** Probleme de sommeil
- hsc17** Sentiment de desesperoir en pensant au futur
- hsc18** Decouragement, morose
- hsc19** Sentiment de solitude
- hsc20** Perte d interets et d envies sexuelles
- hsc21** Sentiment de s etre fait prendre au piège ou fait prisonnier
- hsc22** Agite ou se tracasse beaucoup
- hsc23** Aucun interet pour quoique ce soit
- hsc24** Sentiment que tout est fatiguant
- hsc25** Sentiment d etre inutile

**Details**

In french

---

Intention	<i>Study five of Rogers, T. &amp; Milkman, K. L. (2016). Reminders through association. Psychological Science, 27, 973-986.</i>
-----------	---

---

**Description**

Participants have to answer to many questions (in a 11-page-survey). For 5 questions (indicated by \$\$ at the beginning of the question), they are told that there is a correct answer and that they will earn \$0.06 if they provide this correct answer. At the beginning of the experiment, there are also told that they will earn a \$0.60 bonus if they choose the answer E on the last question (whatever this is the correct answer or not).

**Usage**

data(Intention)

**Format****age** age**choice** Did participants choose to have a reminder? (1 = yes; 0 = no). Note that in conditions 2 and 4, participants had no choices and therefore, 0 is coded for all subjects in these two conditions**Condition** Condition 1 = free-reminder-through-association condition: participants read that they can choose to have (for free) an image of an elephant (presented on screen) that would appear at the bottom of page 11 as a reminder of selecting answer E; Condition 2 = non condition: no reminders; Condition 3 = costly-reminder-through-association condition: participants read that if they pay \$0.03, an image of an elephant (presented on screen) would appear at the bottom of page 11 as a reminder of selecting answer E Condition 4 = forced-reminder-through-association condition: participants read that an image of an elephant (presented on screen) would appear at the bottom of page 11 as a reminder of selecting answer E.**correct** Did participants earn \$0.60 bonus? (1 = yes; 0 = no)**dup** No available information**fee\_for\_reminder** How much was paid for a reminder? (\$0.00 or \$0.03)**filter\_.** No available information**final\_problem** Earned money for answering E on the last question: \$0.00 (if E was not selected) or \$0.60 (if E was selected)**gender** Gender; 0 = male; 1 = female**id** participants id**plus** Earned money at the beginning ( \$0.06 for all participants)**problem1** First question for which participants earn a \$0.03 bonus if they provide the correct answer

**problem2** Second question for which participants earn a \$0.03 bonus if they provide the correct answer

**problem3** Third question for which participants earn a \$0.03 bonus if they provide the correct answer

**problem4** Fourth question for which participants earn a \$0.03 bonus if they provide the correct answer

**problem5** Fifth question for which participants earn a \$0.03 bonus if they provide the correct answer

**Total\_Amount\_Earned**  $\text{Intention}_{\text{final\_problem}}$  minus  $\text{Intention}_{\text{fee\_for\_reminder}}$ ; They are 4 possible outcomes: (1) \$-0.03, if a reminder was paid and answer E was not selected on the last question; (2) \$0.00, if no reminder was paid and answer E was not selected on the last question; (3) \$0.57, if a reminder was paid and answer E was selected on the last question; (4) \$0.60, if no reminder was paid and answer E was selected on the last question

**Total\_Amount\_Earned\_if.forced.to.pay.for.cue** equals  $\text{Intention}_{\text{Total\_Amount\_Earned}}$  in all but one condition: in condition 1 (free-reminder-through-association condition):  $\text{Intention}_{\text{Total\_Amount\_Earned\_if.forced}} - \text{Intention}_{\text{Total\_Amount\_Earned}} - 0.03$

Morality

*Replication of Experiments Evaluating Impact of Psychological Distance on Moral Judgment (Eyal, Liberman & Trope, 2008; Gong & Medin, 2012) Study 2*

## Description

For 6 scenarios, participants have to evaluate the wrongness of actions, with a scale ranging from 1 (not ok) to 5 (completely ok) Contributors: Biljana Jokic Iris Zezelj osf link: <https://osf.io/8wqvc/>

## Usage

```
data(Morality)
```

## Format

a data frame with 145 rows and 10 columns

**number** participant id

**Orig\_rep** Is participant English or Serbian?

**social\_distance** Is the person in the scenario someone participants know (i.e. colleague, neighbor)?

**swing\_r** A girl pushing another kid off a swing because she really wants to use it before going home

**flag\_r** A woman cutting it up a national flag into small pieces and using it in order to clean her house

**hands\_r** A man eating his food with his hands, like most of his family members, also in public, after he washes them

**mother\_r** A loving man who promised her dying mother that he would visit her grave every week but didn't keep his promise because he was very busy

**kiss\_r** Two cousins kissing each other passionately on the mouth, in secret, because there are in love

**dog\_r** Eating our dog that was bitten by a car in front of our house and was killed

**mean\_judge\_r** average of all scenarios judgment

---

outliers\_mad                      *MAD function to detect outliers*

---

## Description

Detecting univariate outliers using the robust median absolute deviation

## Usage

```
outliers_mad(x, b, threshold, na.rm)
```

## Arguments

x	vector of values from which we want to compute outliers
b	constant depending on the assumed distribution underlying the data, that equals $1/Q(0.75)$ . When the normal distribution is assumed, the constant 1.4826 is used (and it makes the MAD and SD of normal distributions comparable).
threshold	the number of MAD considered as a threshold to consider a value an outlier
na.rm	set whether Missing Values should be excluded (na.rm = TRUE) or not (na.rm = FALSE) - defaults to TRUE

## Value

Returns Call, median, MAD, limits of acceptable range of values, number of outliers

## Examples

```
#### Run outliers_mad
x <- runif(150,-100,100)
outliers_mad(x, b = 1.4826,threshold = 3,na.rm = TRUE)

#### Results can be stored in an object.
data(Intention)
res1=outliers_mad(Intention$age)
# Moreover, a list of elements can be extracted from the function,
# such as all the extremely high values,
# That will be sorted in ascending order
#### The function should be performed on dimension rather than on isolated items
data(Attacks)
```

```
SOC <- rowMeans(Attacks[,c("soc1r", "soc2r", "soc3r", "soc4", "soc5", "soc6",  
"soc7r", "soc8", "soc9", "soc10r", "soc11", "soc12", "soc13")])  
res=outliers_mad(x = SOC)
```

---

outliers\_mahalanobis *mahalanobis function to detect outliers*

---

## Description

Detecting multivariate outliers using the Mahalanobis distance

## Usage

```
outliers_mahalanobis(x, alpha, na.rm)
```

## Arguments

x	matrix of bivariate values from which we want to compute outliers
alpha	nominal type I error probability (by default .01)
na.rm	set whether Missing Values should be excluded (na.rm = TRUE) or not (na.rm = FALSE) - defaults to TRUE

## Value

Returns Call, Max distance, number of outliers

## Examples

```
#### Run outliers_mahalanobis  
data(Attacks)  
SOC <- rowMeans(Attacks[,c("soc1r", "soc2r", "soc3r", "soc4", "soc5", "soc6", "soc7r",  
"soc8", "soc9", "soc10r", "soc11", "soc12", "soc13")])  
HSC <- rowMeans(Attacks[,22:46])  
res <- outliers_mahalanobis(x = cbind(SOC,HSC), na.rm = TRUE)  
# A list of elements can be extracted from the function,  
# such as the position of outliers in the dataset  
# and the coordinates of outliers  
res$outliers_pos  
res$outliers_val
```

---

outliers_mcd	<i>MCD function to detect outliers</i>
--------------	--

---

### Description

Detecting multivariate outliers using the Minimum Covariance Determinant approach

### Usage

```
outliers_mcd(x, h, alpha, na.rm)
```

### Arguments

x	matrix of bivariate values from which we want to compute outliers
h	proportion of dataset to use in order to compute sample means and covariances
alpha	nominal type I error probability (by default .01)
na.rm	set whether Missing Values should be excluded (na.rm = TRUE) or not (na.rm = FALSE) - defaults to TRUE

### Value

Returns Call, Max distance, number of outliers

### Examples

```
#### Run outliers_mcd
# The default is to use 75% of the datasets in order to compute sample means and covariances
# This proportion equals 1-breakdown points (i.e. h = .75 <--> breakdown points = .25)
# This breakdown points is encouraged by Leys et al. (2018)
data(Attacks)
SOC <- rowMeans(Attacks[,c("soc1r","soc2r","soc3r","soc4","soc5","soc6","soc7r",
"soc8","soc9","soc10r","soc11","soc12","soc13")])
HSC <- rowMeans(Attacks[,22:46])
res <- outliers_mcd(x = cbind(SOC,HSC), h = .5,na.rm = TRUE)
res

# Moreover, a list of elements can be extracted from the function,
# such as the position of outliers in the dataset
# and the coordinates of outliers
res$outliers_pos
res$outliers_val
```



---

plot\_outliers\_mad      *Plotting function for the mad*

---

**Description**

plotting data and highlighting univariate outliers detected with the outliers\_mad function

**Usage**

```
plot_outliers_mad(res, x, pos_display = FALSE)
```

**Arguments**

res	result of the outliers_mad function from which we want to create a plot
x	data from which the outliers_mad function was performed
pos_display	set whether the position of outliers in the dataset should be displayed on the graph (pos_display = TRUE) or not (pos_display = FALSE)

**Value**

None

**Examples**

```
#### Run outliers_mad and perform plot_outliers_mad on the result
data(Intention)
res=outliers_mad(Intention$age)
plot_outliers_mad(res,x=Intention$age)

### when the number of outliers is small, one can display the outliers position in the dataset
x=c(rnorm(10),3)
res2=outliers_mad(x)
plot_outliers_mad(res2,x,pos_display=TRUE)
```

---

plot\_outliers\_mahalanobis      *Plotting function for the Mahalanobis distance approach*

---

**Description**

plotting data and highlighting multivariate outliers detected with the mahalanobis distance approach

**Usage**

```
plot_outliers_mahalanobis(res, x, pos_display = FALSE)
```

**Arguments**

res	result of the outliers_mad function from which we want to create a plot
x	matrix of bivariate values from which we want to compute outliers
pos_display	set whether the position of outliers in the dataset should be displayed on the graph (pos_display = TRUE) or not (pos_display = FALSE)

**Details**

plotting data and highlighting multivariate outliers detected with the MCD function. Additionally, the plot returns two regression lines: the first one including all data and the second one including all observations but the detected outliers. It allows to observe how much the outliers influence of outliers on the regression line.

**Value**

None

**Examples**

```
#### Run plot_outliers_mahalanobis
data(Attacks)
SOC <- rowMeans(Attacks[,c("soc1r", "soc2r", "soc3r", "soc4", "soc5", "soc6",
"soc7r", "soc8", "soc9", "soc10r", "soc11", "soc12", "soc13")])
HSC <- rowMeans(Attacks[,22:46])
res <- outliers_mahalanobis(x = cbind(SOC,HSC))
plot_outliers_mahalanobis(res, x = cbind(SOC,HSC))

# it's also possible to display the position of the multivariate outliers on the graph
# preferably, when the number of multivariate outliers is not too high
c1 <- c(1,4,3,6,5,2,1,3,2,4,7,3,6,3,4,6)
c2 <- c(1,3,4,6,5,7,1,4,3,7,50,8,8,15,10,6)
res2 <- outliers_mahalanobis(x = cbind(c1,c2))
plot_outliers_mahalanobis(res2, x = cbind(c1,c2),pos_display = TRUE)

# When no outliers are detected, only one regression line is displayed
c3 <- c(1,4,3,6,5)
c4 <- c(1,3,4,6,5)
res3 <- outliers_mahalanobis(x = cbind(c3,c4))
plot_outliers_mahalanobis(res3,x = cbind(c3,c4))
```

**Description**

plotting data and highlighting multivariate outliers detected with the MCD function. Additionally, the plot returns two regression lines: the first one including all data and the second one including all observations but the detected outliers. It allows to observe how much the outliers influence of outliers on the regression line.

**Usage**

```
plot_outliers_mcd(res, x, pos_display = FALSE)
```

**Arguments**

res	result of the outliers_mcd function from which we want to create a plot
x	matrix of bivariate values from which we want to compute outliers
pos_display	set whether the position of outliers in the dataset should be displayed on the graph (pos_display = TRUE) or not (pos_display = FALSE)

**Value**

None

**Examples**

```
#### Run plot_outliers_mcd
data(Attacks)
SOC <- rowMeans(Attacks[,c("soc1r","soc2r","soc3r","soc4","soc5","soc6",
"soc7r","soc8","soc9","soc10r","soc11","soc12","soc13")])
HSC <- rowMeans(Attacks[,22:46])
res <- outliers_mcd(x = cbind(SOC,HSC),na.rm=TRUE)
plot_outliers_mcd(res,x = cbind(SOC,HSC))

# it's also possible to display the position of the multivariate outliers on the graph
# preferably, when the number of multivariate outliers is not too high
c1 <- c(1,4,3,6,5,2,1,3,2,4,7,3,6,3,4,6)
c2 <- c(1,3,4,6,5,7,1,4,3,7,50,8,8,15,10,6)
res2 <- outliers_mcd(x = cbind(c1,c2),na.rm=TRUE)
plot_outliers_mcd(res2, x=cbind(c1,c2),pos_display=TRUE)

# When no outliers are detected, only one regression line is displayed
c3 <- c(1,2,3,1,4,3,6,5)
c4 <- c(1,2,3,1,3,4,6,5)
res3 <- outliers_mcd(x = cbind(c3,c4),na.rm=TRUE)
plot_outliers_mcd(res3,x=cbind(c3,c4),pos_display=TRUE)
```

# Index

- \*Topic **Anxiety**
  - Attacks, [2](#)
  - Morality, [5](#)
- \*Topic **Attacks**
  - Attacks, [2](#)
  - Morality, [5](#)
- \*Topic **Brussels**
  - Attacks, [2](#)
  - Morality, [5](#)
- \*Topic **Coherence**
  - Attacks, [2](#)
  - Morality, [5](#)
- \*Topic **Depression**
  - Attacks, [2](#)
  - Morality, [5](#)
- \*Topic **Intention**
  - Intention, [4](#)
- \*Topic **MAD**
  - outliers\_mad, [6](#)
  - plot\_outliers\_mad, [9](#)
- \*Topic **MMCD**
  - outliers\_mcd, [8](#)
  - plot\_outliers\_mcd, [10](#)
- \*Topic **Reminder,**
  - Intention, [4](#)
- \*Topic **mahalanobis**
  - outliers\_mahalanobis, [7](#)
  - plot\_outliers\_mahalanobis, [9](#)
- \*Topic **outliers**
  - outliers\_mad, [6](#)
  - outliers\_mahalanobis, [7](#)
  - outliers\_mcd, [8](#)
  - plot\_outliers\_mad, [9](#)
  - plot\_outliers\_mahalanobis, [9](#)
  - plot\_outliers\_mcd, [10](#)
- \*Topic **plot**
  - plot\_outliers\_mad, [9](#)
  - plot\_outliers\_mahalanobis, [9](#)
  - plot\_outliers\_mcd, [10](#)
- Attacks, [2](#)
- Intention, [4](#)
- Morality, [5](#)
- outliers\_mad, [6](#)
- outliers\_mahalanobis, [7](#)
- outliers\_mcd, [8](#)
- plot\_outliers\_mad, [9](#)
- plot\_outliers\_mahalanobis, [9](#)
- plot\_outliers\_mcd, [10](#)