

# Package ‘caseMatch’

July 22, 2025

**Type** Package

**Title** Identify Similar Cases for Qualitative Case Studies

**Version** 1.1.0

**Date** 2023-06-05

**Author** Rich Nielsen

**Maintainer** Rich Nielsen <rnielsen@mit.edu>

**Description** Allows users to identify similar cases for qualitative case studies using statistical matching methods.

**License** GPL (>= 2)

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2023-06-07 04:10:02 UTC

## Contents

|                             |   |
|-----------------------------|---|
| caseMatch-package . . . . . | 1 |
| case.match . . . . .        | 2 |
| EU . . . . .                | 5 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>6</b> |
|--------------|----------|

---

|                   |  |
|-------------------|--|
| caseMatch-package | <i>A package for using matching to select cases from a quantitative data set for further qualitative analysis.</i> |
|-------------------|--|

---

## Description

This package uses statistical matching to identify "most similar" cases in a quantitative data set for subsequent qualitative analysis. Unlike existing matching packages, this package intended to meet some specific needs of analysts using matching for case studies.

## Details

Use the `case.match` function.

## Author(s)

Maintainer: Rich Nielsen <rnielsen@mit.edu>

## References

Nielsen, Richard. 2016. "Case Selection via Matching," *Sociological Methods and Research*, 45 (3): 569-597. <http://journals.sagepub.com/doi/abs/10.1177/0049124114547054>

## See Also

[case.match](#)

## Examples

```
data(EU)
mvars <- c("socialist", "rgdpc", "FHc", "FHp", "trade")
dropvars <- c("countryname", "population")

## In this example, I subset to the first 40 obs. to cut run-time
out <- case.match(data=EU[1:40,], id.var="countryname", leaveout.vars=dropvars,
                 distance="mahalanobis", case.N=2, greedy.match="pareto",
                 number.of.matches.to.return=10,
                 treatment.var="eu", max.variance=TRUE)

out$cases
```

---

case.match

*Uses matching methods to select cases for qualitative analysis*

---

## Description

Uses matching methods to select cases for qualitative analysis

## Usage

```
case.match(data, id.var, case.N = 2, distance = "mahalanobis",
           design.type = "most similar", match.case = NULL,
           greedy.match="pareto", number.of.matches.to.return = 1,
           treatment.var = NULL, outcome.var= NULL, leaveout.vars = NULL,
           max.variance = FALSE, max.variance.outcome=FALSE,
           variance.tolerance = 0.1, max.spread = FALSE,
           max.spread.outcome=FALSE, varweights = NULL)
```

**Arguments**

|  |  |
|--|--|
| <code>data</code>                        | A data frame.  |
| <code>id.var</code>                      | A string variable that uniquely identifies cases within the data   |
| <code>case.N</code>                      | The number of cases to choose. Must be 1 or more.  |
| <code>distance</code>                    | The distance metric, specified as a string. Options are "mahalanobis", "euclidean", or "standardized", where "standardized" means that variables are standardized by their standard deviations.  |
| <code>design.type</code>                 | Should the algorithm pick cases that are most similar or most different? Specify either "most similar" or "most different" as a string.  |
| <code>match.case</code>                  | If specified, this is the value of <code>id.var</code> of a specific case to match.  |
| <code>number.of.matches.to.return</code> | How many matches to return.  |
| <code>greedy.match</code>                | Specifies which matches to return. Options are "pareto", "greedy", and "all". "all" keeps all matches. "pareto" matches eliminate 'redundant' matches where both units have better available matches. "greedy" matches keeps only the top matches in the data, but does eliminates best matches for some units since it uses a without replacement algorithm.                                |
| <code>treatment.var</code>               | The name of the treatment variable, specified as a string.   |
| <code>outcome.var</code>                 | The name of the outcome variable, specified as a string.   |
| <code>leaveout.vars</code>               | A vector of variables to not include in the matching.  |
| <code>max.variance</code>                | Should the cases be selected to maximize variance on <code>treatment.var</code> ? If cases should be in opposite treatment conditions, specify <code>max.variance=TRUE</code> .  |
| <code>max.variance.outcome</code>        | Should the cases be selected to maximize variance on <code>outcome.var</code> ? If cases should have opposite outcomes, specify <code>max.variance.outcome=TRUE</code> .   |
| <code>variance.tolerance</code>          | The proportion of cases to consider if <code>max.variance</code> is specified but there are too few cases that maximize the variance of <code>treatment.var</code> .   |
| <code>max.spread</code>                  | Should the cases be selected to maximize "spread" on the treatment variable, meaning that cases are selected to be have evenly values from the min of <code>treatment.var</code> to the max?   |
| <code>max.spread.outcome</code>          | Should the cases be selected to maximize "spread" on the outcome variable, meaning that cases are selected to be have evenly values from the min of <code>outcome.var</code> to the max?   |
| <code>varweights</code>                  | An optional vector of variable weights. It must line up with the columns of the data after <code>id.var</code> and <code>leaveout.vars</code> are removed. Optionally, element names can be included for <code>varweights</code> — if so, the function checks that the names are identical to (and line up with) the names of the matching variables. It will throw an error if they do not. |

**Details**

`case.match` uses statistical matching to select cases in a quantitative data set for subsequent qualitative analysis in "most similar" and "most different" research designs.

**Value**

case.match returns a named list with the following elements:

cases                    A table of the matched cases.  
 case.distances    A list of the distances between matched cases.

**Author(s)**

Rich Nielsen

**References**

Nielsen, Richard. 2016. "Case Selection via Matching," Sociological Methods and Research, 45 (3): 569-597. <http://www.mit.edu/~rnielsen/Case>

**Examples**

```
data(EU)
mvars <- c("socialist","rgdpc","FHc","FHp","trade")
dropvars <- c("countryname","population")

## In this example, I subset to the first 40 obs. to cut run-time
out <- case.match(data=EU[1:40,], id.var="countryname", leaveout.vars=dropvars,
                  distance="mahalanobis", case.N=2,
                  number.of.matches.to.return=10,
                  treatment.var="eu", max.variance=TRUE)

out$cases

## Not run:
## All cases:
## Find the best matches of EU to non-EU countries
out <- case.match(data=EU, id.var="countryname", leaveout.vars=dropvars,
                  distance="mahalanobis", case.N=2,
                  number.of.matches.to.return=10,
                  treatment.var="eu", max.variance=TRUE)

out$cases

## Find the best matches while downweighting political variables
myvarweights <- c(1,1,.1,.1,.1)
names(myvarweights) <- c("rgdpc","trade","FHp","FHc","socialist")
myvarweights
(case.match(data=EU, id.var="countryname", leaveout.vars=dropvars,
            distance="mahalanobis", case.N=2,
            number.of.matches.to.return=10, treatment.var="eu",
            max.variance=TRUE,varweights=myvarweights))$cases

## Find the best non-EU matches for Germany
tabGer <- case.match(data=EU, match.case="German Federal Republic",
                    id.var="countryname",leaveout.vars=dropvars,
                    distance="mahalanobis", case.N=2,
                    number.of.matches.to.return=10,max.variance=TRUE,
                    treatment.var="eu")
```

```
## End(Not run)
```

---

|    |   |
|----|---|
| EU | <i>Cross-national data for 189 countries.</i> |
|----|---|

---

### Description

A cross-national data set including economic and political variables for 189 countries, averaged from 1980-1992.

### Usage

```
data(EU)
```

### Format

A data frame with 185 observations on the following 13 variables.

countryname The name of the country

population Country population from Gleditsch.

rgdpc GDP per capita from Gleditsch.

trade Trade from Gleditsch.

FHp Freedom House political rights.

FHc Freedom House civil rights.

socialist An indicator for countries that were socialist during the Cold War.

eu An indicator for EU members.

### Details

A cross-national data set including economic and political variables for 189 countries, averaged from 1980-1992. Data are collected by Gleditsch and Freedom House.

### Source

Gleditsch, Kristian Skrede. (2004) Expanded Trade and GDP Data, Version 4.0. <http://privatewww.essex.ac.uk/~ksg/exptrade>  
<http://www.freedomhouse.org/report-types/freedom-world>

### References

Nielsen, Richard A. Forthcoming. "Case Selection Via Matching," Sociological Methods and Research. <http://www.mit.edu/~rnielsen/Case Selection via Matching.pdf>

### Examples

```
data(EU)
```

# Index

- \* **~case selection**

- case.match, [2](#)

- \* **~matching**

- case.match, [2](#)

- \* **datasets**

- EU, [5](#)

- \* **package**

- caseMatch-package, [1](#)

case.match, [2](#), [2](#)

caseMatch (caseMatch-package), [1](#)

caseMatch-package, [1](#)

EU, [5](#)