

# Package ‘prcr’

August 25, 2017

**Type** Package

**Title** Person-Centered Analysis

**Version** 0.1.5

**Maintainer** Joshua M Rosenberg <jmichaelrosenberg@gmail.com>

**Description** Provides an easy-to-use yet adaptable set of tools to conduct person-center analysis using a two-step clustering procedure. As described in Bergman and El-Khoury (1999) <DOI:10.1002/(SICI)1521-4036(199910)41:6%3C753::AID-BIMJ753%3E3.0.CO;2-K>, hierarchical clustering is performed to determine the initial partition for the subsequent k-means clustering procedure.

**License** MIT + file LICENSE

**URL** <https://github.com/jrosen48/prcr>

**BugReports** <https://github.com/jrosen48/prcr/issues>

**LazyData** TRUE

**Imports** dplyr, tidyr, ggplot2, tibble, irr, lpSolve, purrr, class

**Suggests** rmarkdown, knitr

**VignetteBuilder** knitr

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Author** Joshua M Rosenberg [aut, cre],  
Jennifer A Schmidt [aut],  
Patrick N Beymer [aut],  
Rebecca R Steingut [ctb]

**Repository** CRAN

**Date/Publication** 2017-08-25 17:20:01 UTC

## R topics documented:

create_profiles . . . . .	2
cross_validate . . . . .	3

detect_outliers . . . . .	4
plot.pcr . . . . .	4
plot_r_squared . . . . .	5
print.pcr . . . . .	6
summary.pcr . . . . .	6

<b>Index</b>	<b>8</b>
--------------	----------

---

create_profiles	<i>Create profiles of observed variables using two-step cluster analysis</i>
-----------------	--

---

## Description

Create profiles of observed variables using two-step cluster analysis

## Usage

```
create_profiles(df, ..., n_profiles, to_center = FALSE, to_scale = FALSE,
  distance_metric = "squared_euclidean", linkage = "complete",
  plot_centered_data = FALSE, plot_raw_data = FALSE)
```

## Arguments

df	with two or more columns with continuous variables
...	unquoted variable names separated by commas
n_profiles	The specified number of profiles to be found for the clustering solution
to_center	Boolean (TRUE or FALSE) for whether to center the raw data with $M = 0$
to_scale	Boolean (TRUE or FALSE) for whether to scale the raw data with $SD = 1$
distance_metric	Distance metric to use for hierarchical clustering; "squared_euclidean" is default but more options are available (see ?hclust)
linkage	Linkage method to use for hierarchical clustering; "complete" is default but more options are available (see ?dist)
plot_centered_data	Boolean (TRUE or FALSE) for whether to center the data before plotting (should not be used if to_center = T; only if to_center = F, in cases in which raw data is used to create profiles but centered profiles are desired for visualization purposes)
plot_raw_data	Boolean (TRUE or FALSE) for whether to plot the raw data, regardless of whether the data are centered or scaled before clustering.

## Details

Function to create a specified number of profiles of observed variables using a two-step (hierarchical and k-means) cluster analysis.

**Value**

A list containing the prepared data, the output from the hierarchical and k-means cluster analysis, the r-squared value, raw clustered data, processed clustered data of cluster centroids, and a ggplot object.

**Examples**

```
create_profiles(mtcars, disp, hp, wt, n_profiles = 2, to_scale = TRUE)
```

---

cross_validate	<i>Returns statistics from double-split cross validation</i>
----------------	--

---

**Description**

Returns statistics from double-split cross validation

**Usage**

```
cross_validate(df, ..., to_center = FALSE, to_scale = FALSE, n_profiles,
              distance_metric = "squared_euclidean", linkage = "complete", k = 30,
              lower_bound = 2, upper_bound = 9)
```

**Arguments**

df	with two or more columns with continuous variables
...	unquoted variable names separated by commas
to_center	(TRUE or FALSE) for whether to center the raw data with $M = 0$
to_scale	Boolean (TRUE or FALSE) for whether to scale the raw data with $SD = 1$
n_profiles	the number of profiles in the solution to cross-validate; or, the character string "iterate" to perform cross-validation for 2 to 9 profile solutions
distance_metric	Distance metric to use for hierarchical clustering; "squared_euclidean" is default but more options are available (see ?hclust)
linkage	Linkage method to use for hierarchical clustering; "complete" is default but more options are available (see ?dist)
k	the number of iterations
lower_bound	if n_profiles = "iterate", then this is the smallest number of profiles in the range of number of profiles to explore; defaults to 2
upper_bound	if n_profiles = "iterate", then this is the largest number of profiles in the range of number of profiles to explore; defaults to 9

**Details**

Performs double-split cross validation and returns Cohen's Kappa and percentage agreement statistics.

**Value**

A ggplot2 object

---

detect_outliers	<i>Identifies potential outliers</i>
-----------------	--------------------------------------

---

**Description**

Identifies potential outliers

**Usage**

```
detect_outliers(df, return_index = TRUE)
```

**Arguments**

df	data.frame (or tibble) with variables to be clustered; all variables must be complete cases
return_index	Boolean (TRUE or FALSE) for whether to return only the row indices of the possible multivariate outliers; if FALSE, then all of the output from the function (including the indices) is returned

**Details**

\* add an argument to 'create\_profiles()' to remove multivariate outliers based on Hadi's (1994) procedure

**Value**

either the row indices of possible multivariate outliers or all of the output from the function, depending on the value of return\_index

---

plot.pcr	<i>Return plot of cluster centroids</i>
----------	---

---

**Description**

Return plot of cluster centroids

**Usage**

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

**Arguments**

x                    A 'prcr' object  
 ...                  Additional arguments

**Details**

Returns ggplot2 plot of cluster centroids

**Value**

A ggplot2 object

---

plot_r_squared	<i>Plot R<sup>2</sup> (r-squared) values for a range of number of profiles</i>
----------------	--

---

**Description**

Plot R<sup>2</sup> (r-squared) values for a range of number of profiles

**Usage**

```
plot_r_squared(df, ..., to_center = FALSE, to_scale = FALSE,
  distance_metric = "squared_euclidean", linkage = "complete",
  lower_bound = 2, upper_bound = 9, r_squared_table = FALSE)
```

**Arguments**

df                    with two or more columns with continuous variables  
 ...                   unquoted variable names separated by commas  
 to\_center            (TRUE or FALSE) for whether to center the raw data with M = 0  
 to\_scale             Boolean (TRUE or FALSE) for whether to scale the raw data with SD = 1  
 distance\_metric     Distance metric to use for hierarchical clustering; "squared\_euclidean" is default but more options are available (see ?hclust)  
 linkage              Linkage method to use for hierarchical clustering; "complete" is default but more options are available (see ?dist)  
 lower\_bound         the smallest number of profiles in the range of number of profiles to explore; defaults to 2  
 upper\_bound         the largest number of profiles in the range of number of profiles to explore; defaults to 9  
 r\_squared\_table     if TRUE, then a table, rather than a plot, is returned; defaults to FALSE

**Details**

Returns ggplot2 plot of cluster centroids

**Value**

A list containing a ggplot2 object and a tibble for the R<sup>2</sup> values

**Examples**

```
df <- mtcars
plot_r_squared(df, mpg, wt, hp, qsec, to_center = TRUE, lower_bound = 2, upper_bound = 4)
```

---

print.pcr	<i>Prints details of pcr cluster solution</i>
-----------	---

---

**Description**

Prints details of pcr cluster solution

**Usage**

```
## S3 method for class 'pcr'
print(x, ...)
```

**Arguments**

x	A 'pcr' object
...	Additional arguments

**Details**

Prints details of of pcr cluster solution

---

summary.pcr	<i>Concise summary of pcr cluster solution</i>
-------------	--

---

**Description**

Concise summary of pcr cluster solution

**Usage**

```
## S3 method for class 'pcr'
summary(object, ...)
```

**Arguments**

object	A 'pcr' object
...	Additional arguments

**Details**

Prints a concise summary of pcr cluster solution

# Index

`create_profiles`, 2

`cross_validate`, 3

`detect_outliers`, 4

`plot.pcr`, 4

`plot_r_squared`, 5

`print.pcr`, 6

`summary.pcr`, 6