Package 'PCovR'

January 20, 2025

Type Package

Title Principal Covariates Regression

Version 2.7.2

Depends GPArotation, ThreeWay, MASS, stats, graphics, Matrix

Description

Analyzing regression data with many and/or highly collinear predictor variables, by simultaneously reducing the predictor variables to a limited number of components and regressing the criterion variables on these components (de Jong S. & Kiers H. A. L. (1992) <doi:10.1016/0169-7439(92)80100-I>). Several rotation and model selection options are provided.

License GPL (>= 2)

Encoding UTF-8

Author Marlies Vervloet [aut, cre],
Henk Kiers [aut],
Eva Ceulemans [ctb]

Maintainer Kristof Meers <kristof.meers+cran@kuleuven.be>

NeedsCompilation no

Repository CRAN

Date/Publication 2023-10-26 07:20:08 UTC

Contents

PCovR-package   *Principal Covariates Regression*

Description

Analyzing regression data with many and/or highly collinear predictor variables, by simultaneously reducing the predictor variables to a limited number of components and regressing the criterion variables on these components. Several rotation options are provided in this package, as well as model selection options.

Details

This package contains the function pcovr, which runs a full PCovR analysis of a data set and provides several preprocessing, model selection, and rotation options. This function calls the function pcovr_est, which estimates the PCovR parameters given a specific weigthing parameter value and a particular number of components. This function was originally written in MATLAB by De Jong & Kiers (1992). Two illustrative data sets are included: alexithymia and psychiatrists.

Author(s)

Marlies Vervloet (<marlies.vervloet@ppw.kuleuven.be>)

References

S. de Jong, H.A.L. Kiers, Principal covariates regression: Part I. Theory, Chemom. intell. lab. syst 14 (1992) 155-164.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

See Also

pcovr

pcovr_est

alexithymia

psychiatrists

Examples

```
data(alexithymia)
results <- pcovr(alexithymia$X, alexithymia$Y)
summary(results)
plot(results)
```

| alexithymia | *Effect of alexithymia on depression and self-esteem* |

Description

The data contain the scores of 122 Belgian psychology students on the 20-item Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994), which measures the inability to recognize and verbalize emotions, the Center for Epidemiological Studies Depression Scale (CES-D; Radloff, 1977), and the Rosenberg Self-Esteem Scale (RSE; Rosenberg, 1989). These data can be used to examine the extent to which the degree of depressive symptomatology (measured by the total CES-D score), and the degree of self-esteem (measured by the total RSE-score), can be predicted by the separate items of the TAS-20. We investigate the individual items because Bankier, Aigner and Bach (2001) emphasize that alexithymia is a multidimensional construct and authors disagree about the number and nature of the dimensions.

Usage

```
data(alexithymia)
```

Format

List of 2

$ X:'data.frame': 122 obs. of 20 variables::

confused I am often confused about what emotion I am feeling

right words It is difficult for me to find the right words for my feelings

sensations I have physical sensations that even doctors don't understand

describe I am able to describe my feelings easily

analyze problems I prefer to analyze problems rather than just describe them

upset When I am upset, I don't know if I am sad, frightened, or angry

puzzled I am often puzzled by sensations in my body

let happen I prefer to just let things happen rather than to understand why they turned out that way

let happen I have feelings that I can't quite identify

essential Being in touch with emotions is essential

feel about people I find it hard to describe how I feel about people

describe more People tell me to describe my feelings more

going on I don't know what's going on inside me

why angry I often don't know why I am angry

daily activities I prefer talking to people about their daily activities rather than their feelings

entertainment I prefer to watch "light" entertainment shows rather than psychological dramas

reveal feelings It is difficult for me to reveal my innermost feelings, even to close friends

close I can feel close to someone, even in moments of silence

useful I find examination of my feelings useful in solving personal problems

hidden meanings Looking for hidden meanings in movies or plays distracts from their enjoyment

$ Y:'data.frame': 122 obs. of 2 variables::

CES-D Degree of depressive symptomatology
RSE Degree of self-esteem

References

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale: Item selection and cross-validation of the factor structure. Journal of Psychosomatic Research , 38 (1), 23-32.

Bankier, B., Aigner, M., & Bach, M. (2001). Alexithymia in DSM-IV Disorder: Comparative Evaluation of Somatoform Disorder, Panic Disorder, Obsessive-Compulsive Disorder, and Depression. Psychosomatics , 42 (3), 235-240.

Radloff, R. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. Applied psychological measurement , 1 (3), 385-401.

Rosenberg, M. (1989). Society and the adolescent self-image. Middletown: Wesleyan University Press.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Examples

```
data(alexithymia)
str(alexithymia)
```

ErrorRatio   *Error variance ratio*

Description

Estimating the ratio of the error variance of the predictor block, versus the error variance of the criterion block.

Usage

```
ErrorRatio(X, Y, Rmin = 1, Rmax = ncol(X)/3, prepX="stand",prepY="stand")
```

Arguments

X	Dataframe containing predictor scores
Y	Dataframe containing criterion scores
Rmin	Lowest number of components considered
Rmax	Highest number of components considered
prepX	Preprocessing of predictor scores: standardizing (stand) or centering data (cent)
prepY	Preprocessing of criterion scores: standardizing (stand) or centering data (cent)

Details

An estimate for the error variance of X can be obtained by applying principal component analysis to X and determining the optimal number of components through a scree test; the estimate equals the associated percentage of unexplained variance. The estimate for the error variance of Y boils down to the percentage of unexplained variance when Y is regressed on X. This approach for estimating and was based on the work of Wilderjans, Ceulemans, Van Mechelen, and Van den Berg (2011).

Value

The returned value is the estimated error variance of X, divided by the estimated error variance of Y.

Author(s)

Marlies Vervloet

References

Wilderjans, T. F., Ceulemans, E., Van Mechelen, I., & Van den Berg, R. A. (2011). Simultaneous analysis of coupled data matrices subject to different amounts of noise. British Journal of Mathematical and Statistical Psychology , 64, 277-290.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Examples

```
data(psychiatrists)
ratio <- ErrorRatio(psychiatrists$X,psychiatrists$Y)
```

---

pcovr   *Full Principal covariates regression analysis of a specific data set*

Description

Application of a PCovR analysis consists of the following steps: preprocessing the data, running PCovR analyses with different numbers of components and/or weighting parameter values, performing model selection, and rotating the retained solution for easier interpretation.

Usage

```
pcovr(X,Y,modsel="seq",Rmin=1,Rmax=ncol(X)/3,R=NULL,weight=NULL,rot="varimax",
target=NULL, prepX="stand",prepY="stand", ratio="estimation", fold="LeaveOneOut",
zeroloads=ncol(X))

## S3 method for class 'pcovr'
plot(x,cpal=NULL,lpal=NULL,...)
```

Arguments

X	Dataframe containing predictor scores
Y	Dataframe containing criterion scores
modsel	Model selection procedure (seq, seqRcv, seqAcv or sim)
Rmin	Lowest number of components considered
Rmax	Highest number of components considered
R	Number of components (overrules Rmin and Rmax)
weight	Weighting values considered
rot	Rotation criterion (varimax, quartimin, targetT, targetQ, wvarim or promin)
target	Target matrix for target rotation (components x predictor variables)
prepX	Preprocessing of predictor scores: standardizing (stand) or centering data (cent)
prepY	Preprocessing of criterion scores: standardizing (stand) or centering data (cent)
ratio	Ratio of the estimated error variances of the predictor block and the criterion block
fold	Value of *k* when performing *k*-fold cross-validation. By default, leave-one-out cross-validation is performed.
zeroloads	Number of near-zero loadings of the target for simplimax rotation
x	An object of the type produced by pcovr
cpal	Vector of colors used for model selection plots
lpal	Vector of line types used for model selection plots
...	Further graphical arguments

Details

Preprocessing: The PCovR package includes two preprocessing options, which can be applied to X and/or Y. Specifically, it is possible to only center the data (prepX="cent", prepY="cent"). However, the default option is to standardize the data (prepX="stand", prepY="stand"), which implies that X and/or Y are centered and normalized (i.e., each variable has a mean of zero and a standard deviation of one).

Model selection:

Sequential procedure: The fastest and therefore default model selection setting (modsel="seq") implies a sequential procedure in which the weighting value is determined on the basis of maximum likelihood principles (Vervloet, Van den Noortgate, Van Deun, & Ceulemans, 2013), but taking the weighting values entered by the user (i.e., specified with the parameter weight) into account. Specifically, if the weighting value does not equal one of those values, the entered weighting value that is closest to the maximum likelihood weighting value (in absolute sense) is used. Note that the default error variance ratio is estimated with the function ErrorRatio, but can be specified otherwise with the parameter ratio. However, this is only possible for datasets with more observations than predictor variables. Among all models with the selected weighting value and a number of components between Rmin and Rmax, the solution is picked that has the highest *st* value (Cattell, 1966; Wilderjans, Ceulemans, & Meers, 2012). However, models for which the fit is less than 1% better than the fit of a less complex model are excluded.

Note that the assessment of the optimal number of components can be overruled, in case one is only interested in the solutions with a particular number of components. In particular, when specifying the input parameter R, Rmin and Rmax will be ignored, and the specified number of components will be used when running the analysis and determining the weighting value.

The package also provides two sequential procedures that incorporate a cross-validation step (modsel="seqRcv" and modsel="seqAcv"). seqRcv also starts with the selection of the weighting value based on maximum likelihood principles, but in the next step, the number of components is determined using leave-one-out cross-validation. seqAcv is identical to the default procedure, but has an extra step: after the selection of the number of components, leave-one-out cross-validation is applied to choose the weighting value.

Simultaneous procedure: The simultaneous procedure (modsel="sim") performs leave-one-out cross-validation for all considered weighting values (weight; by default, 100 values between .01 and 1) and all numbers of components between Rmin (default: 1) and Rmax (default: number of predictors divided by 3). The weighting parameter value and number of components that maximize the cross-validation fit are retained. Note that the parameter fold can be used to alter the number of roughly equal-sized parts in which the data are split for cross-validation (Hastie, Tibshirani, & Friedman, 2001). The default value of fold is "LeaveOneOut", implying that the data is split in *N* (number of observations) parts.

Interpreting the component matrices: The rotation criteria that are implemented in the PCovR package are varimax, quartimin, targetT, targetQ, wvarim and promin. One can also request the original solution by typing rot="none". Target rotation (Browne, 1972) orthogonally rotates the loading matrix towards a target matrix (target) that is specified by the user. Note that Simplimax requires the specification of a number of zero elements. By default, zeroloads equals the number of predictors.

The interpretation of the obtained solution usually starts with the interpretation of the loading matrix. Specifically, the components are labeled by considering what the predictors that have the highest loadings (in absolute sense), have in common. Given these labels, the regression weights can be interpreted.

Value

pcovr returns a list that contains the following objects (note that some objects can be empty, depending on the model selection settings used) :

Px	Loading matrix (components x predictor variables)
Py	Regression weights matrix (components x criterion variables)
Te	Component score matrix (observations x components)
W	Component weights matrix (predictor variables x components)
Rx2	Proportion of explained variance in X
Ry2	Proportion of explained variance in Y
Qy2	Cross-validation fit as a function of weighting parameter and number of components (weighting parameter x number of components)
VAFsum	Weighted sum of the variance accounted for in X and in Y as a function of number of components (1 x number of components)
alpha	Selected value of the weighting parameter

R	Selected number of components
modsel	Model selection procedure that was used
rot	Rotation criterion that was used
prepX	Method of preprocessing that was used for the predictor scores
prepY	Method of preprocessing that was used for the criterion scores
Rvalues	Numbers of components that were considered
Alphavalues	Weighting parameter values that were considered

Author(s)

Marlies Vervloet (<marlies.vervloet@ppw.kuleuven.be>)

References

Browne, M. W. (1972). Oblique rotation to a partially specified target. British Journal of Mathematical and Statistical Psychology , 25 (2), 207-212.

Cattell, R. B. (1966). The scree test for the number of factors. Multivariate behavioral research , 1 (2), 245-276.

De Jong, S., & Kiers, H. A. (1992). Principal covariates regression: Part I. Theory. Chemometrics and Intelligent Laboratory Systems , 155-164.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning: Data mining, inference and prediction. New York: Springer.

Vervloet, M., Van Deun, K., Van den Noortgate, W., & Ceulemans, E. (2013). On the selection of the weighting parameter value in Principal Covariates Regression. Chemometrics and Intelligent Laboratory Systems.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Wilderjans, T. F., Ceulemans, E., & Meers, K. (2012). CHull: A generic convex-hull-based model selection method. Behavior research methods .

Examples

```
data(alexithymia)
results <- pcovr(alexithymia$X, alexithymia$Y)
summary(results)
plot(results)
```

---

pcovr_est	*Estimation of Principal Covariates Regression parameters, given a prespecified weighting value and number of components*

Description

Analyzing regression data with many and/or highly collinear predictor variables, by simultaneously reducing the predictor variables to a given number of components and regressing the criterion variables on these components. A weighting parameter value is specified that determines the extent to which both aspects influence the solution. Cross-validation (Hastie, Tibshirani & Friedman, 2001) options are included.

Usage

```
pcovr_est(X, Y, r, a, cross = FALSE, fold = "LeaveOneOut")
```

Arguments

X	Matrix containing predictor scores (observations x predictors)
Y	Matrix containing criterion scores (observations x criteria)
r	The desired number of components
a	The desired weighting parameter value
cross	Logical. If TRUE cross-validation is performed
fold	Value of *k* when performing *k*-fold cross-validation. By default, leave-one-out cross-validation is performed.

Value

W	Component weights matrix (predictors x components)
B	Regression weights for predictors (predictors x criteria)
Rx2	Proportion of explained variance in X
Ry2	Proportion of explained variance in Y
Te	Component score matrix (observations x components)
Px	Loading matrix of components (components x predictors)
Py	Regression weights matrix (components x criteria)
Qy2	Cross-validation fit

Author(s)

Marlies Vervloet (<marlies.vervloet@ppw.kuleuven.be>)

References

De Jong, S., & Kiers, H. A. (1992). Principal covariates regression: Part I. Theory. Chemometrics and Intelligent Laboratory Systems , 155-164.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). The elements of statistical learning: Data mining, inference and prediction. New York: Springer.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Examples

```
data(alexithymia)
X <- data.matrix(alexithymia$X)
Y <- data.matrix(alexithymia$Y)
results <- pcovr_est(X, Y, r=2, a=.90)
str(results)
```

---

promin | *Promin rotation*

Description

This is a rotation criterion, developed by Lorenzo-Seva (1999), in which oblique target rotation (`tarrotob`) is applied using the Weighted Varimax solution (`wvarim`) as the target matrix.

Usage

```
promin(F1, nrs = 20)
```

Arguments

F1	Matrix to be rotated
nrs	Number of random starts

Value

Th	Transformation matrix to the pattern
loadings	Rotated matrix
U	Transformation matrix to the structure

Author(s)

Marlies Vervloet (<marlies.vervloet@ppw.kuleuven.be>)

References

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. Multivariate Behavioral Research, 34(3), 347-365.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Examples

```
Px <- matrix(rnorm(36),12,3)
print(Px)

Px_r <- promin(Px)
print(Px_r$loadings)
```

---

psychiatrists	*Effect of psychiatric symptoms on toxicomania, schizophrenia, depression and anxiety disorder*

Description

The data contain the scores of 30 Belgian psychiatric patients on 23 psychiatric symptoms and 4 psychiatric disorders (toxicomania, schizophrenia, depression, and anxiety disorder). Each score is the sumscore of the binary symptom and disorder scores that were given by 15 different psychiatrists. The data can be used to examine the extent to which the degree of toxicomania, schizophrenia, depression and anxiety disorder, can be predicted by the 23 psychiatric symptoms.

Usage

```
data(psychiatrists)
```

Format

The format is: List of 2

$ X:'data.frame': 30 obs. of 23 variables::

```
desorganised_speech
agitation
hallucinations
inappropriate
desorientation
depression
fear
suicide
somatic_concern
```

```
narcotics
antisocial
retardation
social_isolation
routine
alcohol
negativism
denial
grandeur
suspicion
intellectual_obstruction
impulse_control
social_leveling
occupational_dysfunction
```

$ Y:'data.frame': 30 obs. of 4 variables::

```
toxicomania
schizophrenia
depression
anxiety_disorder
```

References

Van Mechelen, I., & De Boeck, P. (1990). Projection of a binary criterion into a model of hierarchical classes. Psychometrika , 55 (4), 677-694.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Examples

```
data(psychiatrists)
str(psychiatrists)
```

SortLoadings   *Sorting a component loading matrix*

Description

A loading matrix indicates how predictors that have been reduced to components - e.g., in principal covariates regression (De Jong & Kiers, 1992) - relate to these components. Usually, components are interpreted by looking at what the predictors with a clear non-zero loading have in common. To make this easier, this function changes the order of the predictors presented in a loading matrix, so that the firstly, the predictors with clear non-zero loadings on the first component (with decreasing loadings) are presented, then the predictors with clear non-zero loadings on the second component, etc.

Usage

```
SortLoadings(Px)
```

Arguments

Px   Dataframe that contains component loadings (components x predictors)

Value

`SortLoadings` returns a dataframe with the same dimensions and labels as the original loading matrix, but with the columns (referring to the predictors) presented in a different order.

Author(s)

Marlies Vervloet (`<marlies.vervloet@ppw.kuleuven.be>`)

References

De Jong, S., & Kiers, H. A. (1992). Principal covariates regression: Part I. Theory. Chemometrics and Intelligent Laboratory Systems , 155-164.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

See Also

pcovr

Examples

```
# Compute loading matrix of alexithymia dataset
data(alexithymia)
results <- pcovr(alexithymia$X,alexithymia$Y)
Px <- results$Px
print(Px)

# Sort loading matrix
sorted_Px <- SortLoadings(results$Px)
print(sorted_Px)
```

---

tarrotob   *Oblique target rotation*

Description

Oblique target rotation

Usage

```
tarrotob(F1, W)
```

Arguments

F1	Matrix to be rotated
W	Target binary matrix

Value

T	Rotation matrix
A	Rotated matrix

Author(s)

Marlies Vervloet (<marlies.vervloet@ppw.kuleuven.be>)

Examples

```
Px <- matrix(rnorm(36),12,3)
print(Px)

W <- matrix(rbinom(36,1,.4),12,3)
Px_r <- tarrotob(Px,W)
print(Px_r$A)
```

---

wvarim   *Weighted varimax*

Description

This is an oblique rotation criterion, developed by Cureton and Mulaik (1975).

Usage

```
wvarim(F1, nrs = 20)
```

Arguments

F1	Matrix to be rotated
nrs	Number of random starts

Value

Th	Rotation matrix
loadings	Rotated matrix
W	Matrix of weights
fr	Varimax function value
ir	Number of iterations

Author(s)

Marlies Vervloet (<marlies.vervloet@ppw.kuleuven.be>)

References

Cureton, E. E., & Mulaik, S. A. (1975). The weighted varimax rotation and the promax rotation. Psychometrika, 40(2), 183-195.

Marlies Vervloet, Henk A. Kiers, Wim Van den Noortgate, Eva Ceulemans (2015). PCovR: An R Package for Principal Covariates Regression. Journal of Statistical Software, 65(8), 1-14. URL http://www.jstatsoft.org/v65/i08/.

Examples

```
Px <- matrix(rnorm(36),12,3)
print(Px)

Px_r <- wvarim(Px)
print(Px_r$loadings)
```

Index