Package ‘PCovR’

February 19, 2015

Type  Package
Title  Principal Covariates Regression
Version  2.5
Depends  GPArotation, ThreeWay, MASS, Matrix
Date  2014-07-14
Description  Analyzing regression data with many and/or highly collinear predictor variables, by simultane-fully 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.
License  GPL (>= 2)
Author  Marlies Vervloet [aut, cre],
          Henk Kiers [aut],
          Eva Ceulemans [ctb]
Maintainer  Marlies Vervloet <marlies.vervloet@ppw.kuleuven.be>
NeedsCompilation  no
Repository  CRAN
Date/Publication  2014-07-16 10:57:35

R topics documented:

PCovR-package .................................................. 2
alexithymia .................................................... 3
ErrorRatio ...................................................... 4
pcovr ........................................................... 5
pcovr_est ...................................................... 8
promin .......................................................... 10
psychiatrists ................................................. 11
SortLoadings .................................................. 12
tarrotob ...................................................... 13
wvarim ......................................................... 14

Index  15
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

Package: **PCovR**
Type: Package
Version: 2.4
Date: 2014-07-14
License: GPL (>= 2)

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 weighting 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.ervloet@ppw.kuleuven.be>)

References


See Also

pcovr
pcovr_est
alexithymia
psychiatrists

Examples

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

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
  have feelings 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
errorratio

Error Ratio

1. I find examination of my feelings useful in solving personal problems.
2. 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


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


Examples

data(psychiatrists)
ratio <- errorratio(psychiatrists$X, psychiatrists$Y)

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=ErrorRatio(X,Y,Rmin,Rmax,prepX,prepY), fold="LeaveOneOut",zeroloads=ncol(X))

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

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>Dataframe containing predictor scores</td>
</tr>
<tr>
<td>y</td>
<td>Dataframe containing criterion scores</td>
</tr>
<tr>
<td>modsel</td>
<td>Model selection procedure (seq, seqRcv, seqAcv or sim)</td>
</tr>
<tr>
<td>rmin</td>
<td>Lowest number of components considered</td>
</tr>
<tr>
<td>rmax</td>
<td>Highest number of components considered</td>
</tr>
<tr>
<td>r</td>
<td>Number of components (overrides rmin and rmax)</td>
</tr>
<tr>
<td>weight</td>
<td>Weighting values considered</td>
</tr>
<tr>
<td>rot</td>
<td>Rotation criterion (varimax, quartimin, targetT, targetQ, wvarim or promin)</td>
</tr>
<tr>
<td>target</td>
<td>Target matrix for target rotation (components x predictor variables)</td>
</tr>
<tr>
<td>prepX</td>
<td>Preprocessing of predictor scores: standardizing (stand) or centering data (cent)</td>
</tr>
<tr>
<td>prepY</td>
<td>Preprocessing of criterion scores: standardizing (stand) or centering data (cent)</td>
</tr>
<tr>
<td>ratio</td>
<td>Ratio of the estimated error variances of the predictor block and the criterion block</td>
</tr>
<tr>
<td>fold</td>
<td>Value of k when performing k-fold cross-validation. By default, leave-one-out cross-validation is performed.</td>
</tr>
<tr>
<td>zeroloads</td>
<td>Number of near-zero loadings of the target for simplimax rotation</td>
</tr>
<tr>
<td>x</td>
<td>An object of the type produced by pcovr</td>
</tr>
<tr>
<td>cpal</td>
<td>Vector of colors used for model selection plots</td>
</tr>
<tr>
<td>lpal</td>
<td>Vector of line types used for model selection plots</td>
</tr>
</tbody>
</table>

**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. 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). 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):

- **P**x: Loading matrix (components x predictor variables)
- **P**y: Regression weights matrix (components x criterion variables)
- **T**e: Component score matrix (observations x components)
- **W**: Component weights matrix (predictor variables x components)
- **R**x2: Proportion of explained variance in X
- **R**y2: Proportion of explained variance in Y
- **Q**y2: 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
Alpha values Weighting parameter values that were considered

Author(s)

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

References


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.
pcovr_est

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


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 (tarroto) is applied using the Weighted Varimax solution (wvarim) as the target matrix.

**Usage**

```r
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.ervloet@ppw.kuleuven.be>)

**References**


**Examples**

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

px_r <- promin(px)
print(px_r$loadings)
```
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(psihiatrists)

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


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>)
tarrotob

References


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

```r
px <- matrix(rnorm(SV), 12, 3)
print(px)

w <- matrix(rbinom(SVLQLNT), 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**

```r
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**


**Examples**

```r
px <- matrix(rnorm(SV), 12, 3)
print(px)

px_r <- wvarim(px)
print(px_r$loadings)
```
Index

*Topic datasets
  alexithymia, 3
  psychiatrists, 11
*Topic package
  PCovR-package, 2
*Topic regression
  ErrorRatio, 4
  pcovr, 5
  pcovr_est, 8
  SortLoadings, 12
*Topic rotation
  promin, 10
  tarroto, 13
  wvarim, 14

alexithymia, 2, 3

colors, 6

ErrorRatio, 4, 6

PCovR (PCovR-package), 2
pcovr, 2, 5, 13
PCovR-package, 2
pcovr_est, 2, 8
plot.pcovr (pcovr), 5
print.pcovr (pcovr), 5
promin, 6, 7, 10
psychiatrists, 2, 11

SortLoadings, 12
summary.pcovr (pcovr), 5

tarroto, 10, 13

wvarim, 6, 7, 10, 14