Package: pcev (via r-universe)

October 16, 2024

Title Principal Component of Explained Variance

Version 2.2.2

Description Principal component of explained variance (PCEV) is a statistical tool for the analysis of a multivariate response vector. It is a dimension- reduction technique, similar to Principal component analysis (PCA), that seeks to maximize the proportion of variance (in the response vector) being explained by a set of covariates.

Depends R (>= 3.0.0)

Imports RMTstat, stats, corpcor

License GPL (>=2)

LazyData true

URL `http://github.com/GreenwoodLab/pcev`

BugReports `http://github.com/GreenwoodLab/pcev/issues`

Suggests knitr, testthat

VignetteBuilder knitr

RoxygenNote 6.1.0

Repository https://greenwoodlab.r-universe.dev

RemoteUrl https://github.com/greenwoodlab/pcev

RemoteRef HEAD

RemoteSha f54217f17f6154f4ac61459e6095215460e50c2e

Contents

`pcev-package` *pcev: A package for computing principal components of explained variance.*

Description

PCEV is a statistical tool for the analysis of a multivariate response vector. It is a dimension-reduction technique, similar to Principal Components Analysis (PCA), which seeks the maximize the proportion of variance (in the response vector) being explained by a set of covariates.

pcev functions

`estimatePcev` `computePCEV` `PcevObj` `permutePval` `wilksPval` `roysPval`

`computePCEV` *Principal Component of Explained Variance*

Description

`computePCEV` computes the first PCEV and tests its significance.

Usage

```
computePCEV(response, covariate, confounder, estimation = c("all",
  "block", "singular"), inference = c("exact", "permutation"),
  index = "adaptive", shrink = FALSE, nperm = 1000,
  na_action = "fail", Wilks = FALSE)
```

Arguments

`response`	A matrix of response variables.
`covariate`	An array or a data frame of covariates.
`confounder`	An array or data frame of confounders.
`estimation`	Character string specifying which estimation method to use: `"all"`, `"block"` or `"singular"`. Default value is `"all"`.
`inference`	Character string specifying which inference method to use: `"exact"` or `"permutation"`. Default value is `"exact"`.
`index`	Only used if `estimation = "block"`. Default value is `"adapative"`. See details.
`shrink`	Should we use a shrinkage estimate of the residual variance? Default value is `FALSE`.
`nperm`	The number of permutations to perform if `inference = "permutation"` or for the Tracy-Widom empirical estimate (if `estimation = "singular"`).

`na_action`	how NAs are treated. The default is to raise an error. See details.
`Wilks`	Should we use a Wilks test instead of Roy's largest test? This is only implemented for a single covariate and with `estimation = "all"`.

Details

This is the main function. It computes the PCEV using either the classical method, block approach or singular. A p-value is also computed, testing the significance of the PCEV.

The p-value is computed using either a permutation approach or an exact test. The implemented exact tests use Wilks' Lambda (only for a single covariate) or Roy's Largest Root. The latter uses Johnstone's approximation to the null distribution. Note that for the block approach, only p-values obtained from a permutation procedure are available.

When `estimation = "singular"`, the p-value is computed using a heuristic: using the method of moments and a small number of permutations (i.e. 25), a location-scale family of the Tracy-Widom distribution of order 1 is fitted to the null distribution. This fitted distribution is then used to compute p-values.

When `estimation = "block"`, there are three different ways of specifying the blocks: 1) if `index` is a vector of the same length as the number of columns in `response`, then it is used to match each response to a block. 2) If `index` is a single positive integer, it is understood as the number of blocks, and each response is matched to a block randomly. 3) If `index = "adaptive"` (the default), the number of blocks is chosen so that there are about n/2 responses per block, and each response is match to a block randomly. All other values of `index` should result in an error.

By default, missing values are not allowed. This can be relaxed with `na_action`. If `na_action = "omit"`, then all rows with at least one missing value will be removed from `response` before computation. If `na_action = "column"`, then the estimation of the linear model parameters is done column-wise with the non-missing value. This approach maximises the information. Note that missing values are still not allowed in `covariate` and `confounder`.

Value

An object of class `Pcev` containing the first PCEV, the p-value, the estimate of the shrinkage factor, etc.

See Also

`estimatePcev`

Examples

```
set.seed(12345)
Y <- matrix(rnorm(100*20), nrow=100)
X <- rnorm(100)
pcev_out <- computePCEV(Y, X)
pcev_out2 <- computePCEV(Y, X, shrink = TRUE)
```

`estimatePcev` *Estimation of PCEV*

Description

`estimatePcev` estimates the PCEV.

Usage

```
estimatePcev(pcevObj, ...)

## Default S3 method:
estimatePcev(pcevObj, ...)

## S3 method for class 'PcevClassical'
estimatePcev(pcevObj, shrink, index, ...)

## S3 method for class 'PcevBlock'
estimatePcev(pcevObj, shrink, index, ...)

## S3 method for class 'PcevSingular'
estimatePcev(pcevObj, shrink, index, ...)
```

Arguments

`pcevObj`	A pcev object of class `PcevClassical`, `PcevBlock` or `PcevSingular`
`...`	Extra parameters.
`shrink`	Should we use a shrinkage estimate of the residual variance?
`index`	If `pcevObj` is of class `PcevBlock`, `index` is a vector describing the block to which individual response variables correspond.

Value

A list containing the variance components, the first PCEV, the eigenvalues of $V_R^{-1}V_M$ and the estimate of the shrinkage parameter ρ

See Also

`computePCEV`

`methylation` *Methylation values around BLK gene*

Description

A dataset containing methylation values for cell-separated samples. The methylation was measured using bisulfite sequencing. The data also contains the genomic position of these CpG sites, as well as a binary phenotype (i.e. whether the sample comes from a B cell).

Usage

```
methylation

pheno

position

index

pheno2

position2

methylation2
```

Format

The data comes in four objects:

methylation Matrix of methylation values at 5,986 sites measured on 40 samples
pheno Vector of phenotype, indicating whether the sample comes from a B cell
position Data frame recording the position of each CpG site along the chromosome
index Index vector used in the computation of PCEV-block
methylation2 Matrix of methylation values at 1000 sites measured on 40 samples
pheno2 Vector of phenotype, indicating the cell type of the sample (B cell, T cell, or Monocyte)
position2 Data frame recording the position of each CpG site along the chromosome

Details

Methylation was first measured at 24,068 sites, on 40 samples. Filtering was performed to keep the 25% most variable sites. See the vignette for more detail.

A second sample of the methylation dataset was extracted. This second dataset contains methylation values at 1000 CpG dinucleotides.

Source

Tomi Pastinen, McGill University, Genome Quebec.

`PcevObj` *Constructor functions for the different pcev objects*

Description

`PcevClassical`, `PcevBlock` and `PcevSingular` create the pcev objects from the provided data that are necessary to compute the PCEV according to the user's parameters.

Usage

```
PcevClassical(response, covariate, confounder)

PcevBlock(response, covariate, confounder)

PcevSingular(response, covariate, confounder)
```

Arguments

`response`	A matrix of response variables.
`covariate`	A matrix or a data frame of covariates.
`confounder`	A matrix or data frame of confounders

Value

A pcev object, of the class that corresponds to the estimation method. These objects are lists that contain the data necessary for computation.

See Also

`estimatePcev`, `computePCEV`

`permutePval` *Permutation p-value*

Description

Computes a p-value using a permutation procedure.

Usage

```
permutePval(pcevObj, ...)

## Default S3 method:
permutePval(pcevObj, ...)

## S3 method for class 'PcevClassical'
permutePval(pcevObj, shrink, index, nperm, ...)

## S3 method for class 'PcevBlock'
permutePval(pcevObj, shrink, index, nperm, ...)

## S3 method for class 'PcevSingular'
permutePval(pcevObj, shrink, index, nperm, ...)
```

Arguments

`pcevObj`	A pcev object of class `PcevClassical` or `PcevSingular` `PcevBlock`
`...`	Extra parameters.
`shrink`	Should we use a shrinkage estimate of the residual variance?
`index`	If `pcevObj` is of class `PcevBlock`, `index` is a vector describing the block to which individual response variables correspond.
`nperm`	The number of permutations to perform.

`roysPval` *Roy's largest root exact test*

Description

In the classical domain of PCEV applicability this function uses Johnstone's approximation to the null distribution of ' Roy's Largest Root statistic. It uses a location-scale variant of the Tracy-Widom distribution of order 1.

Usage

```
roysPval(pcevObj, ...)

## Default S3 method:
roysPval(pcevObj, ...)

## S3 method for class 'PcevClassical'
roysPval(pcevObj, shrink, index, ...)

## S3 method for class 'PcevSingular'
roysPval(pcevObj, shrink, index, nperm, ...)
```

```
## S3 method for class 'PcevBlock'
roysPval(pcevObj, shrink, index, ...)
```

Arguments

`pcevObj`	A pcev object of class `PcevClassical` or `PcevBlock`
`...`	Extra parameters.
`shrink`	Should we use a shrinkage estimate of the residual variance?
`index`	If `pcevObj` is of class `PcevBlock`, `index` is a vector describing the block to which individual response variables correspond
`nperm`	Number of permutations for Tracy-Widom empirical estimate.

Details

Note that if `shrink` is set to `TRUE`, the location-scale parameters are estimated using a small number of permutations.

`wilksPval` *Wilks' lambda exact test*

Description

Computes a p-value using Wilks' Lambda.

Usage

```
wilksPval(pcevObj, ...)

## Default S3 method:
wilksPval(pcevObj, ...)

## S3 method for class 'PcevClassical'
wilksPval(pcevObj, shrink, index, ...)

## S3 method for class 'PcevSingular'
wilksPval(pcevObj, shrink, index, ...)

## S3 method for class 'PcevBlock'
wilksPval(pcevObj, shrink, index, ...)
```

Arguments

`pcevObj`	A pcev object of class `PcevClassical` or `PcevBlock`
`...`	Extra parameters.
`shrink`	Should we use a shrinkage estimate of the residual variance?
`index`	If `pcevObj` is of class `PcevBlock`, `index` is a vector describing the block to which individual response variables correspond.

Details

The null distribution of this test statistic is only known in the case of a single covariate, and therefore this is the only case implemented.

Index