Package ‘profileR’

February 20, 2015

Title Profile Analysis of Multivariate Data in R

Type Package

Description The package profileR provides a set of multivariate methods and data visualization tools to implement profile analysis and cross-validation techniques described in Davison & Davenport (2002) and Bulut (2013).

Version 0.2-1

Date 12-20-2013

Author Okan Bulut <okanbulut84@gmail.com>, Christopher David Desjardins <cddesjardins@gmail.com>

Maintainer Christopher David Desjardins <cddesjardins@gmail.com>

Depends ggplot2, MASS, RColorBrewer, reshape, methods, plyr, R (>= 3.0.0)

License GPL (>= 2)

LazyData true

Collate 'anova.critpat.R' 'cp.R' 'pams.R' 'pbg.R' 'pc.R'
   'plot.critpat.R' 'plot.prof.R' 'pr.R' 'print.critpat.R'
   'print.prof.R' 'print.profg.R' 'profileplot.R'
   'summary.critpat.R' 'summary.profg.R'

NeedsCompilation no

Repository CRAN

Date/Publication 2013-12-20 23:14:29

R topics documented:

profileR-package ................................................. 2
cp ................................................................. 2
EEGS ............................................................. 4
IPMMc ............................................................ 5
leisure .......................................................... 6
pams ............................................................... 6
pbg ............................................................... 7
Profile Analysis of Multivariate Data in R

Description

The package profileR provides a set of multivariate methods and data visualization tools to implement profile analysis and cross-validation techniques described in Davison & Davenport (2002) and Bulut (2013).

Author(s)

Okan Bulut <okanbulut84@gmail.com>
Christopher David Desjardins <cddesjardins@gmail.com>

References


Criterion-Related Profile Analysis

Description

Implements the criterion-related profile analysis described in Davison & Davenport (2002).
Usage

\texttt{cp(criterion, predictor, k = 100, na.action = "na.fail")}

\texttt{\# S3 method for class 'critpat'}
\texttt{summary(object, ...)}

\texttt{\# S3 method for class 'critpat'}
\texttt{plot(x, ...)}

\texttt{\# S3 method for class 'critpat'}
\texttt{anova(object, ...)}

\texttt{\# S3 method for class 'critpat'}
\texttt{print(x, ...)}

Arguments

criterion \hspace{1cm} \text{Corresponds to the criterion variable.}
predictor \hspace{1cm} \text{Corresponds to the predictor variables.}
k \hspace{1cm} \text{Corresponds to the scalar constant and must be greater than 0. Defaults to 100.}
na.action \hspace{1cm} \text{How should missing data be handled? Function defaults to failing if missing data are present.}
x \hspace{1cm} \text{An object of class "critpat".}
object \hspace{1cm} \text{An object of class "critpat".}
... \hspace{1cm} \text{Additional arguments to be passed.}

Details

The \texttt{critpat} function requires two arguments: criterion and predictors. The function returns the criterion-related profile analysis described in Davison & Davenport (2002). Missing data are presently handled by specifying \texttt{na.action = "na.omit"}, which performs listwise deletion and \texttt{na.action = "na.fail"}, the default, which causes the function to fail. The following s3 generic functions are available: \texttt{summary()}, \texttt{anova()}, \texttt{print()}, and \texttt{plot()}. These functions provide a summary of the analysis (namely, R2 and the level and pattern components); perform ANOVA of the R2 for the pattern, the level, and the overall model; provide output similar to \texttt{lm()}, and plots the pattern effect.

Value

An object of class \texttt{critpat} is returned, listing the following components:

\texttt{lvl.comp} \hspace{1cm} \text{The level component}
\texttt{pat.comp} \hspace{1cm} \text{The pattern component}
b \hspace{1cm} \text{The regression weights}
bstar \hspace{1cm} \text{The mean centered regression weights}
xc \hspace{1cm} \text{The scalar constant times bstar}
Description

The EEGS is a subset of the Entrance Examination for Graduate Studies in Turkey. There are three subscores in EEGS: Quantitative 1, Quantitative 2, and Verbal. In order to show the utility of subscore reliability method in this package, each subtest was separated into two parallel form.
Description

The IPMMc data frame has 6 rows and 5 columns. See Davison & Davenport (2002) for more information.

Format

This data frame contains the following columns:

- A Anxiety
- H Hypochondriasis
- S Schizophrenia
- B Bipolar Disorder
- R The Neurotic versus Psychotic Criterion Variable, where Neurotic (= 1) or Psychotic (= 0)

Source


References

Leisure Activity Rankings

Description

The leisure dataset includes leisure activity rankings for three different groups: politicians, administrators, and belly-dancers. Rankings are provided in three categories: Reading-TV, Dance-TV, and TV-Ski. See Tabachnik & Fidell (1996) for more details.

Source


Examples

```r
## Not run:
data(leisure)
## End(Not run)
```

Profile Analysis via Multidimensional Scaling

Description

The pams function implements profile analysis via multidimensional scaling as described by Davison, Davenport, & Bielinski (1995) and Davenport, Ding, & Davison (1995).

Usage

```r
pams(x, dim)
```

Arguments

- `x`: A data matrix or data frame; rows represent individuals, columns represent scores; missing scores are not allowed.
- `dim`: Number of dimensions to be extracted from the data.

Details

The pams function computes similarity/dissimilarity indices based on Euclidean distances between the scores provided in the data, and then extracts dimensional coordinates for each score using multidimensional scaling. A weight matrix, level parameters, and fit measures are computed for each subject in the data.
**Value**

- **dimensional.configuration**
  A matrix that provides prototypical profiles of dimensions extracted from the data

- **weights.matrix**
  A matrix that includes the subject correspondence weights for all dimensions, level parameters, and the subject fit measure which is the proportion of variance in the subject’s actual profiles accounted for by the prototypical profiles.

**Author(s)**

Okan Bulut <okanbulut84@gmail.com>

**References**


**See Also**

cp, pr

**Examples**

```r
## Not run:
data(PS)
result <- pams(PS[,2:4], dim=2)
result

## End(Not run)
```

---

**pbg**

*Profile by group: Testing Parallel, Coincidental, and Level Profiles*

**Description**

The `pbg` function implements three hypothesis tests. These tests are whether the profiles are parallel, coincidental, and level across two groups defined by the grouping variable. If parallelism is rejected, the other two tests are meaningless. In that case, flatness may be assessed within each group, and various within- and between-group contrasts may be analyzed.

**Usage**

```r
pbg(x, y, original.names=FALSE, profile.plot=FALSE, ...)
```

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

```r
```
Arguments

x A matrix or data frame with multiple scores; rows represent individuals, columns represent subscores. Missing subscores have to be inserted as NA.

y A vector or data frame that indicates a grouping variable. It can be either numeric or character (e.g., male-female, high-low, 0-1). The grouping variable must have the same length of x. Missing values are not allowed in y.

original.names Use original column names in x. If FALSE, variables are renamed using v1, v2, ... for subscores and "group" for the grouping variable. Default is FALSE.

profile.plot Print a profile plot of scores for two groups. Default is FALSE.

object An object of class "profg".

Value

An object of class profg is returned, listing the following components:

data.summary Means of observed variables by the grouping variable

corr.table A matrix of correlations among observed variables splitted by the grouping variable

profile.test Results of F-tests for testing parallel, coincidential, and level profiles across two groups.

Author(s)

Okan Bulut <okanbulut84@gmail.com>

See Also

pr, profileplot

Examples

## Not run:
#Read spouse data from Professor Richard Jonhson's website
spouse <- read.table(file="http://www.stat.wisc.edu/~rich/JWMULT02dat/T6-14.DAT")
mod <- pbg(spouse[,1:4],spouse[,5],labels=FALSE,profile.plot=TRUE)
print(mod) #prints average scores in the profile across two groups
summary(mod) #prints the results of hypothesis tests
**Description**

Implements the cross-validation described in Davison & Davenport (2002).

**Usage**

```r
pc(criterion, predictor, seed = NULL, na.action = "na.fail")
```

**Arguments**

- `criterion`: Corresponds to the criterion variable
- `predictor`: Corresponds to the predictor variable
- `seed`: Should a seed be set? Function defaults to a random seed.
- `na.action`: How should missing data be handled? Function defaults to failing if missing data are present.

**Details**

The `pc` function requires two arguments: criterion and predictor. The criterion corresponds to the dependent variable and the predictor corresponds to the matrix of predictor variables. The function performs the cross-validation technique described in Davison & Davenport (2002) and an object of class `critpat` is returned. There the following s3 generic functions are available: `summary()`, `anova()`, `print()`, and `plot()`. These functions provide a summary of the cross-validation (namely, R2); performs ANOVA of the R2 based on the split for the level, pattern, and overall; provide output similar to `lm()`; and plot the estimated parameters for the random split. Missing data are presently handled by specifying `na.action = "na.omit"`, which performs listwise deletion and `na.action = "na.fail"`, the default, which causes the function to fail. A seed may also be set for reproducibility by setting the seed.

**Value**

- `R2.full`: Test of the null hypothesis that $R^2 = 0$
- `R2.pat`: Test that the $R^2_{\text{pattern}} = 0$
- `R2.level`: Test that the $R^2_{\text{level}} = 0$
- `R2.full.lvl`: Test that the $R^2_{\text{full}} = R^2_{\text{level}} = 0$
- `R2.full.pat`: Test that the $R^2_{\text{full}} = R^2_{\text{pattern}} = 0$

**Author(s)**

Christopher David Desjardins <cddesjardins@gmail.com>
References


See Also

cp, print.critpat, summary.critpat, anova.critpat, plot.critpat

---

**pr**  
*Pattern and Level Reliability via Profile Analysis*

**Description**

The *pr* function uses subscores from two parallel test forms and computes profile reliability coefficients as described in Bulut (2013).

**Usage**

```r
pr(form1, form2)
```

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

**Arguments**

- `form1, form2`: Two data matrices or data frames; rows represent individuals, columns represent subscores. Both forms should have the same individuals and subscores in the same order. Missing subscores have to be inserted as NA.
- `x`: An object of class "prof" returned from the *pr* function.
- `...`: Additional arguments to be passed.

**Details**

Profile pattern and level reliability coefficients are based on the profile analysis approach described in Davison and Davenport (2002) and Bulut (2013). Using the parallel test forms or multiple administration of the same test form, pattern and level reliability coefficients are computed. Pattern reliability is an indicator of variability between the subscores of an examinee and the level reliability is an indicator of the average subscore variation among all examinees. For details, see Bulut (2003).

**Value**

An object of class prof is returned, listing the following components:

- `reliability`: Within-person, between-person, and overall subscore reliability
- `pattern.level`: A matrix of all pattern and level values obtained from the subscores
profileplot

Author(s)

Okan Bulut <okanbulut84@gmail.com>

References


See Also

profileplot

Examples

```r
## Not run:
data(eegs)
result <- pr(eegs[,c(1,3,5)],eegs[,c(2,4,6)])
print(result)
plot(result)
```

profileplot | Score Profile Plot

Description

The `profileplot` function creates a profile plot for a matrix or dataframe with multiple scores or subscores using `ggplot2` function in `ggplot2` package.

Usage

```r
profileplot(form,person.id,standardize=TRUE,interval=10,
by.pattern=TRUE,original.names=TRUE)
```

Arguments

- **form**: A matrix or dataframe including two or more subscores.
- **person.id**: A vector that includes person ID values (Optional).
- **standardize**: If not FALSE, all scores are rescaled with a mean of 0 and standard deviation of 1. Default is TRUE.
- **interval**: The number of equal intervals from the mimimum score to the maximum score. Default is 10. Ignored when by.pattern=FALSE.
by.pattern If TRUE, the function creates a profile plot with level and pattern values using ggplot2. Otherwise, the function creates a profile plot showing profile scores of persons using the base graphics in R. Default is TRUE.

original.names Use the original column names in the data. Otherwise, columns are renamed as v1,v2,... Default is TRUE.

Value

The profileplot functions returns a score profile plot from either ggplot2 or the base graphics in R.

Author(s)

Okan Bulut <okanbulut84@gmail.com>

See Also

ggplot2, PS

Examples

```r
## Not run:
data(PS)
myplot <- profileplot(PS[,2:4],person.id=PS$Person,by.pattern=TRUE,original.names=TRUE)
myplot
data(leisure)
leis.plot <- profileplot(leisure[,2:4],standardize=TRUE,by.pattern=FALSE)
leis.plot
## End(Not run)
```

---

**PS**  
*A Hypothetical Personality Scale from Davison, Kim, & Close (2009)*

Description

The PS shows score profiles of six respondents to a hypothetical personality scale. It includes three types of profile patterns: Linearly increasing, inverted V, and linearly decreasing.

Format

- **Person**  Person ID
- **NEU**  Neurotic scale score
- **PSY**  Psychotic scale score
- **CD**  Character disorder scale score
Source


References

Index

*Topic **datasets**
  EEKS, 4
  IPMMc, 5
  leisure, 6
  PS, 12

*Topic **methods**
  cp, 2
  pc, 9

*Topic **package**
  profileR-package, 2

anova.critpat, 10
anova.critpat(cp), 2

cp, 2, 7, 10

EEKS, 4

ggplot2, 11, 12

IPMMc, 5

leisure, 6

pams, 6
pbg, 7
pc, 4, 9
plot.critpat, 10
plot.critpat(cp), 2
plot.prof(pr), 10
pr, 7, 8, 10
print.critpat, 10
print.critpat(cp), 2
profileplot, 8, 11, 11
profileR-package, 2
PS, 12, 12

summary.critpat, 10
summary.critpat(cp), 2
summary.profg(pbg), 7