Package 'powerPLS'

March 21, 2024

Type Package

Title Power Analysis for PLS Classification

Version 0.1.0

Description It estimates power and sample size for Partial Least Squares-based methods described in Andreella, et al., (2024) <arXiv:2403.10289>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.1

Imports compositions, ks, mvtnorm, simukde, nipals, MASS, foreach

Language en-US

BugReports https://github.com/angeella/powerPLS/issues

URL https://github.com/angeella/powerPLS

Depends R (>= 2.10)

NeedsCompilation no

Author Angela Andreella [aut, cre] (Main author,
<https://orcid.org/0000-0002-1141-3041>)

Maintainer Angela Andreella <angela.andreella@unive.it>

Repository CRAN

Date/Publication 2024-03-21 16:40:02 UTC

R topics documented:

Index   **11**

---

computePower   *Power estimation*

Description

estimate power for a given sample size, alpha level and number of score components.

Usage

```
computePower(X, Y, A, n, seed = 123,
Nsim = 100, nperm = 200, alpha = 0.05,
test = "R2", Y.prob = FALSE, eps = 0.01, ...)
```

Arguments

X	data matrix where columns represent the p variables and rows the n observations.
Y	data matrix where columns represent the two classes and rows the n observations.
A	number of score components
n	sample size
seed	seed value
Nsim	number of simulations
nperm	number of permutations
alpha	type I error
test	type of test, one of c("score", "mcc", "R2").
Y.prob	Boolean value. Default @FALSE. IF @TRUE Y is a probability vector
eps	Default 0.01. eps is used when Y.prob = FALSE to transform Y in a probability vector. Default to "R2".
...	Futher parameters see [PLSc](#)

Value

Returns the corresponding estimated power

Author(s)

Angela Andreella

References

Andreella, A., Finos, L., Scarpa, B. and Stocchero, M. "Towards a power analysis for PLS-based methods" arXiv:2403.10289 stat.ME.

Examples

```
datas <- simulatePilotData(nvar = 10, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- computePower(X = datas$X, Y = datas$Y, A = 3, n = 20)
```

---

computeSampleSize	*sample size estimation*

Description

Compute optimal sample size calculation

Usage

```
computeSampleSize(n, X, Y, A, alpha, beta,
nperm, Nsim, seed, test = "R2",...)
```

Arguments

n	vector of sample sizes to consider
X	data matrix where columns represent the p variables and rows the n observations.
Y	data matrix where columns represent the two classes and rows the n observations.
A	number of score components
alpha	level of type I error. Default 0.05
beta	level of type II error. Default 0.2.
nperm	number of permutations. Default 100.
Nsim	number of simulations. Default 100.
seed	seed value
test	type of test, one of c("score", "mcc", "R2"). Default to @R2.
...	Futher parameters.

Value

Returns a dataframe that contains the estimated power for each sample size and number of components considered

Author(s)

Angela Andreella

References

Andreella, A., Finos, L., Scarpa, B. and Stocchero, M. "Towards a power analysis for PLS-based methods" arXiv:2403.10289 stat.ME.

Examples

```
datas <- simulatePilotData(nvar = 10, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- computeSampleSize(X = datas$X, Y = datas$Y, A = 2, n = 20)
```

---

mccTest	*MCC test*

Description

Performs randomization test based on Matthews Correlation Coefficient

Usage

```
mccTest(X, Y, nperm = 100, A, randomization = FALSE, Y.prob = FALSE, eps = 0.01,...)
```

Arguments

X	data matrix where columns represent the p variables and rows the n observations.
Y	data matrix where columns represent the two classes and rows the n observations.
nperm	number of permutations. Default 100.
A	number of score components
randomization	Boolean value. Default @FALSE. If @TRUE the permutation p-value is computed
Y.prob	Boolean value. Default @FALSE. IF @TRUE Y is a probability vector
eps	Default 0.01. eps is used when Y.prob = FALSE to transform Y in a probability vector
...	Futher parameters.

Value

Returns a list with the corresponding statistical tests, raw and adjusted p-values

List with the following objects: pv: raw p-value, pv_adj: adjusted p-value, test estimated statistical test.

Author(s)

Angela Andreella

References

For the general framework of power analysis for PLS-based methods see:

See Also

The type of tests implemented: scoreTest R2Test.

Examples

```
datas <- simulatePilotData(nvar = 30, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- mccTest(X = datas$X, Y = datas$Y, A = 1)
out
```

PLSc   *PLS*

Description

Performs Partial Least Squares two class

Usage

```
PLSc(X, Y, A, scaling = "auto-scaling", post.transformation = TRUE,
eps = 0.01, Y.prob = FALSE, transformation = "ilr")
```

Arguments

X	data matrix where columns represent the p variables and rows the n observations.
Y	data matrix where columns represent the two classes and rows the n observations.
A	number of score components
scaling	type of scaling, one of c("auto-scaling", "pareto-scaling", "mean-centering"). Default @auto-scaling
post.transformation	
	Boolean value. @TRUE if you want to apply post transformation. Default @TRUE
eps	Default 0.01. eps is used when Y.prob = FALSE to transform Y in a probability vector
Y.prob	Boolean value. Default @FALSE. IF @TRUE Y is a probability vector
transformation	transformation used to map Y in probability data vector. The options are @ilr and @clr. Default @ilr.

Value

Returns a list with the following objects: - `W`: matrix of weights - `X_loading`: matrix of X loading - `Y_loading`: matrix of Y loading - `X`: matrix of X data - `Y`: matrix of Y data - `T_score`: matrix of scores - `Y_fitted`: fitted Y matrix - `B`: Matrix regression coefficients - `M`: number of orthogonal components if post transformation is applied.

Author(s)

Angela Andreella

References

Stocchero, M., De Nardi, M., & Scarpa, B. (2021). PLS for classification. Chemometrics and Intelligent Laboratory Systems, 216, 104374.

Examples

```
datas <- simulatePilotData(nvar = 30, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- PLSc(X = datas$X, Y = datas$Y, A = 3)
```

R2Test	*R2 test*

Description

Performs randomization test based on R2

Usage

```
R2Test(X, Y, nperm = 100, A, randomization = FALSE, Y.prob = FALSE, eps = 0.01,...)
```

Arguments

X	data matrix where columns represent the p variables and rows the n observations.
Y	data matrix where columns represent the two classes and rows the n observations.
nperm	number of permutations. Default 100.
A	number of score components
randomization	Boolean value. Default @FALSE. If @TRUE the permutation p-value is computed
Y.prob	Boolean value. Default @FALSE. IF @TRUE Y is a probability vector
eps	Default 0.01. eps is used when Y.prob = FALSE to transform Y in a probability vector
...	Futher parameters.

Value

Returns a list with the corresponding statistical tests, raw and adjusted p-values

List with the following objects: pv: raw p-value, pv_adj: adjusted p-value, test estimated statistical test.

Author(s)

Angela Andreella

References

For the general framework of power analysis for PLS-based methods see:

See Also

The type of tests implemented: scoreTest mccTest.

Examples

```
datas <- simulatePilotData(nvar = 30, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- R2Test(X = datas$X, Y = datas$Y, A = 1)
out
```

---

scoreTest	*Score test*

Description

Performs randomization test based on predictive score vector

Usage

```
scoreTest(X, Y, nperm = 100, A, randomization = FALSE, Y.prob = FALSE, eps = 0.01,...)
```

Arguments

X	data matrix where columns represent the p variables and rows the n observations.
Y	data matrix where columns represent the two classes and rows the n observations.
nperm	number of permutations. Default 100.
A	number of score components
randomization	Boolean value. Default @FALSE. If @TRUE the permutation p-value is computed
Y.prob	Boolean value. Default @FALSE. IF @TRUE Y is a probability vector
eps	Default 0.01. eps is used when Y.prob = FALSE to transform Y in a probability vector
...	Further parameters.

Value

Returns a list with the corresponding statistical tests, raw and adjusted p-values

List with the following objects: pv: raw p-value, pv_adj: adjusted p-value, test estimated statistical test.

Author(s)

Angela Andreella

References

For the general framework of power analysis for PLS-based methods see:

See Also

The type of tests implemented: mccTest R2Test.

Examples

```
datas <- simulatePilotData(nvar = 30, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- scoreTest(X = datas$X, Y = datas$Y, A = 1)
out
```

---

simulatePilotData   *simulate pilot data*

Description

simulate cluster pilot data

Usage

```
simulatePilotData(seed = 123, nvar, clus.size, nvar_rel,m, A = 2, S1 = NULL, S2 = NULL)
```

Arguments

seed	seed value
nvar	number of variables
clus.size	Vector of two elements, specifying the size of classes (only two classes are considered)
nvar_rel	number of variables relevant to predict Y
m	separation between classes
A	oracle number of score components
S1	covariance matrix for the first class. Default @NULL i.e., the identity is considered.
S2	covariance matrix for the second class. Default @NULL i.e., the identity is considered.

Value

Returns list of X and Y simulated data

Author(s)

Angela Andreella

Examples

```
datas <- simulatePilotData(nvar = 10, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
```

---

sim_XY   *simulate data matrix under the alternative hypothesis*

Description

simulate data matrix under the alternative hypothesis

Usage

```
sim_XY(out, n, seed = 123, post.transformation = TRUE, A)
```

Arguments

out	output from PLSc
n	number of observations to simulate
seed	seed value
post.transformation	
	Boolean value. Default @TRUE i.e., post transformation is applied.
A	number of score components used in PLSc.

Value

Returns a simulated matrix under the alternative hypothesis.

Author(s)

Angela Andreella

Examples

```
datas <- simulatePilotData(nvar = 10, clus.size = c(5,5),m = 6,nvar_rel = 5,A = 2)
out <- PLSc(X = datas$X, Y = datas$Y, A = 2)
out_sim <- sim_XY(out = out, n = 10, A = 2)
```

---

umor_acqueo	*This is data to be included in my package*

Description

This is data to be included in my package

Usage

```
umor_acqueo
```

Format

An object of class tbl_df (inherits from tbl, data.frame) with 59 rows and 45 columns.

Author(s)

Angela Andreella <angela.andreella@unive.it>

References

https://pubmed.ncbi.nlm.nih.gov/31069551/

Index