

Package ‘SCOUTer’

June 22, 2020

Title Simulate Controlled Outliers

Version 0.0.0.9000

Description

SCOUTer offers a new way of simulating outliers. Using PCA as a base model, the user can generate new observations th

License file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.0

Imports readxl,
ggplot2,
ggpubr,
xlsx,
stats

R topics documented:

barwithucl	2
custombar	2
distplot	3
distplotsimple	4
ht2info	5
obscontribpanel	5
pcamb_classic	6
pcame	6
scoreplot	7
scoreplotsimple	8
scout	9
scoutgrid	10
scoutsimple	11
scoutsteps	11
speinfo	12
xshift	13

Index	14
--------------	-----------

barwithucl	<i>Single bar plot with Upper Control Limis. Customized title and labels. Y-Axis limits are fixed according to the range of the values in x.</i>
------------	--

Description

Single bar plot with Upper Control Limis. Customized title and labels. Y-Axis limits are fixed according to the range of the values in x.

Usage

```
barwithucl(
  x,
  iobs,
  ucl,
  plotname = "",
  ylabelname = "",
  xlabelname = "Obs. Index"
)
```

Arguments

x	vector with the values of the statistic.
iobs	index of the observations whose value will be displayed.
ucl	Upper Control Limit of the statistic.
plotname	string with the title of the plot. Set to "" by default.
ylabelname	string with the y-axis label. Set to "" by default.
xlabelname	string with the y-axis label. Set to "Obs. Index" by default.

Value

ggplot object with the individual value of a variable as a geom_col with an horizontal line reference.

custombar	<i>Bar plot with with customized title and labels. Y-Axis limits are fixed according to the range of the values in X.</i>
-----------	---

Description

Bar plot with with customized title and labels. Y-Axis limits are fixed according to the range of the values in X.

Usage

```
custombar(
  X,
  iobs,
  plotname = "",
  ylabelname = "Contribution",
  xlabelname = "ggplot2:."
)
```

Arguments

<code>X</code>	matrix with observations as row vectors.
<code>iobs</code>	index of the observations whose value will be displayed.
<code>plotname</code>	string with the title of the plot. Set to "" by default.
<code>ylabelname</code>	string with the y-axis label. Set to "Contribution" by default.
<code>xlabelname</code>	string with the y-axis label. Set to "Variables" by default.

Value

ggplot object with the values of a vector with a customized `geom_col` layer.

<code>distplot</code>	<i>Returns the distance plot according to the input arguments. Coordinates are expressed in terms of the Hotelling's T^2 (x-axis) and the Squared Prediction Error (y-axis) obtained projecting X on the provided pca model. Observations can be identified by the <code>obstag</code> input argument.</i>
-----------------------	--

Description

Returns the distance plot according to the input arguments. Coordinates are expressed in terms of the Hotelling's T^2 (x-axis) and the Squared Prediction Error (y-axis) obtained projecting X on the provided pca model. Observations can be identified by the `obstag` input argument.

Usage

```
distplot(
  X,
  pcaref,
  obstag = matrix(0, nrow(X), 1),
  plottitle = "Distance plot\n"
)
```

Arguments

<code>X</code>	data matrix with observations to be displayed in the distance plot.
<code>pcaref</code>	list with the information of the PCA model.
<code>obstag</code>	Optional column vector of integers indicating the group of each observation (0 or 1). Default value set to <code>matrix(0, nrow(X), 1)</code> .
<code>plottitle</code>	Optional string with the plot title. Set to "Distance plot" by default.

Value

ggplot object with the distance plot.

distplotsimple

Distance plot

Description

Returns the distance plot according to the input arguments. Coordinates are expressed in terms of the Hotelling's T^2 (x-axis) and the Squared Prediction Error (y-axis) obtained projecting X on the provided pca model. Observations can be identified by the obstag input argument.

Usage

```
distplotsimple(  
  T2,  
  SPE,  
  lim.t2,  
  lim.spe,  
  ncomp,  
  obstag = matrix(0, length(T2), 1),  
  alpha = 0.05,  
  plottitle = "Distance plot\n"  
)
```

Arguments

T2	Vector with the Hotelling's T^2 values for each observation.
SPE	Vector with the SPE values for each observation.
lim.t2	Value of the Upper Control Limit for the T^2 statistic.
lim.spe	Value of the Upper Control Limit for the SPE.
ncomp	An integer indicating the number of PCs.
obstag	Optional column vector of integers indicating the group of each observation (0 or 1). Default value set to <code>matrix(0, nrow(X), 1)</code> .
alpha	Optional number between 0 and 1 expressing the type I risk assumed in the computation of the Upper Control Limits set to 0.05 (5 %) by default.
plottitle	Optional string with the plot title, "Distance plot" by default.

Value

distplotobj ggplot object with the generated distance plot.

ht2info	<i>Returns information about T^2 statistic for an observation. Two subplots show the information of an observation regarding its T^2 statistic, i.e.: a bar plot indicating the value of the statistic for the observation, and a bar plot with the contribution that each component had for the T^2 value</i>
---------	---

Description

Returns information about T^2 statistic for an observation. Two subplots show the information of an observation regarding its T^2 statistic, i.e.: a bar plot indicating the value of the statistic for the observation, and a bar plot with the contribution that each component had for the T^2 value

Usage

```
ht2info(HT2, T2matrix, limht2, iobs = NA)
```

Arguments

HT2	A vector with values of the Hotelling's T^2_A statistic.
T2matrix	A matrix with the contributions of each PC (A columns) for each observation (rows) to the Hotelling's T^2_A statistic.
limht2	Upper Control Limit for the Hotelling's T^2_A statistic, at a certain confidence level $(1-\alpha)*100$ %.
iobs	Integer with the index of the observation of interest. Default value set to NA.

Value

ggplot object with the generated bar plots.

obscontribpanel	<i>Information about T^2 and SPE statistics of an observation.</i>
-----------------	---

Description

Information about T^2 and SPE statistics of an observation.

Usage

```
obscontribpanel(pcax, pcaref, obsid = NA)
```

Arguments

pcax	A list with the elements of the PCA model that will be displayed: SPE, T^2_A and their contributions (E and T2matrix).
pcaref	A list with the PCA model according to which the distance and contributions are expressed.
obsid	Integer with the index of the observation of interest. Default set to NA.

Value

ggplot object with the generated bar plots in a 1 x 4 subplots layout.

pcamb_classic	<i>PCA model fitting according to a matrix X using svd.</i>
---------------	---

Description

PCA model fitting according to a matrix X using svd.

Usage

```
pcamb_classic(X, ncomp, alpha, prepro)
```

Arguments

X	Matrix with observations that will used to fit the PCA model.
ncomp	An integer indicating the number of PCs that the model will have.
alpha	A number between 0 and 1 indicating the type I risk assumed to calculate the Upper Control Limits for the SPE, the T ² _A and the scores. The confidence level of these limits will be (1-alpha)*100.
prepro	A string indicating the preprocessing to be performed on X. Its possible values are: "none", for any preprocessing, "cent", for a mean-centering, or "autosc", for a mean-centering and unitary variance scaling (autoscaling).

Value

list with elements containing information about PCA model: m (mean vector), s (standard deviation vector), P (loading matrix with the loadings of each PC stored as columns), Pfull (full loading matrix obtained by the svd), lambda (vector with the variance of each PC), limspe (Upper Control Limit for the SPE with a confidence level (1-alpha)*100 %), limt2 (Upper Control Limit for the T²_A with a confidence level (1-alpha)*100 %), limits_t (Upper control Limits for the scores with a confidence level (1-alpha)*100 %), prepro (string indicating the type of preprocessing performed on X), ncomp (number of PCs of the PCA model, A), alpha (value of the type I risk assumed to calculate the Upper Control Limits of the SPE, T²_A and scores), n (dimension of the number of rows in X), S (covariance matrix of X).

pcame	<i>Projection of X onto a PCA model.</i>
-------	--

Description

pcame performs the projection of the data in X onto the PCA model stored as a list of parameters. It returns the projection of the observations in X, along with the SPE, Hotelling's T²_A, contribution elements and the reconstruction of X obtained by the PCA model.

Usage

```
pcame(X, pcaref)
```

Arguments

<code>X</code>	Matrix with observations that will be projected onto the PCA model.
<code>pcaref</code>	A list with the elements of a PCA model: <code>m</code> (mean), <code>s</code> (standard deviation), <code>prepro</code> (preprocessing: "none", "cent" or "autosc"), <code>P</code> (loading matrix), <code>lambda</code> (vector with variances of each PC).

Value

list with elements containing information about `X` in the PCA model: `Xpreprocessed` (matrix `X` preprocessed), `Tscores` (score matrix with the projection of `X` on each one of the `A` PCs), `E` (error matrix with the par of `X` not explained by the PCA model), `SPE` (vector with the SPE for each observation of `X`), `T2` (vector with the T^2_A for each observation of `X`), `T2matrix` (matrix with the contributions of each PC to the T^2_A for each observation of `X`) and `Xrec` (matrix with the reconstructed part of `X`, i.e. the part of `X` explained by the PCA model).

<code>scoreplot</code>	<i>Score plot and confidence ellipsoid according to the provided pca model. Includes the option of distinguishing between observations and a customized title.</i>
------------------------	--

Description

Score plot and confidence ellipsoid according to the provided `pca` model. Includes the option of distinguishing between observations and a customized title.

Usage

```
scoreplot(
  X,
  pcamodel,
  obstag = matrix(0, nrow(X), 1),
  pcx = 1,
  pcy = 2,
  alpha = 0.05,
  plottitle = "Score plot\n"
)
```

Arguments

<code>X</code>	Matrix with the data to be displayed.
<code>pcamodel</code>	List with the PCA model elements.
<code>obstag</code>	Optional column vector of integers indicating the group of each observation (0 or 1). Default value set to <code>matrix(0, nrow(X), 1)</code> .
<code>pcx</code>	Optional integer with the number of the PC in the horizontal axis. Set to 1 by default.
<code>pcy</code>	Optional integer with the number of the PC in the vertical axis. Set to 2 by default.
<code>alpha</code>	Optional number between 0 and 1 expressing the type I risk assumed in the computation of the confidence ellipse, set to 0.05 (5 %) by default.
<code>plottitle</code>	Optional string with the plot title. Set to "Score plot" by default.

Value

ggplot object with the generated score plot.

scoreplotsimple	<i>Score plot and confidence ellipsoid according to the provided pca model. Includes the option of distinguishing between observations and a customized title.</i>
-----------------	--

Description

Score plot and confidence ellipsoid according to the provided pca model. Includes the option of distinguishing between observations and a customized title.

Usage

```
scoreplotsimple(
  Tscores,
  pcx = 1,
  pcy = 2,
  obstag = matrix(0, nrow(Tscores), 1),
  alpha = 0.05,
  varT = stats::var(Tscores),
  plottitle = "Score plot\n"
)
```

Arguments

Tscores	Matrix with the scores to be displayed, with the information of each PC stored by columns.
pcx	Optional integer with the number of the PC in the horizontal axis. Set to 1 by default.
pcy	Optional integer with the number of the PC in the vertical axis. Set to 2 by default.
obstag	Optional column vector of integers indicating the group of each observation (0 or 1). Default value set to <code>matrix(0, nrow(X), 1)</code> .
alpha	Optional number between 0 and 1 expressing the type I risk assumed in the computation of the confidence ellipse, set to 0.05 (5 %) by default.
varT	Optional parameter expressing the variance of each PC. Set to <code>var(Tscores)</code> by default.
plottitle	Optional string with the plot title. Set to "Score plot" by default.

Value

ggplot object with the generated score plot.

scout

*Shift of an observation following a selected pattern.***Description**

Shift of an observation following a selected pattern.

Usage

```
scout(
  X,
  pcaref,
  T2.y = NA,
  SPE.y = NA,
  nsteps = 1,
  nsteps.spe = 1,
  nsteps.t2 = 1,
  gspe = 1,
  gt2 = 1,
  mode = "simple"
)
```

Arguments

X	Matrix with observations that will be shifted as rows.
pcaref	List with the elements of a PCA model: m (mean), s (standard deviation), prepro (preprocessing: "none", "cent" or "autosc"), P (loading matrix), lambda (vector with variances of each PC).
T2.y	A number indicating the target value for the T^2_A after the shift. Set to NA by default.
SPE.y	A number indicating the target value for the SPE after the shift. Set to NA by default.
nsteps	A number indicating the number of steps between the reference and target values of the SPE and the T^2 . Set to 1 by default.
nsteps.spe	An integer indicating the number of steps in which the shift from the reference to the target value of the SPE will be performed. Set to 1 by default
nsteps.t2	An integer indicating the number of steps in which the shift from the reference to the target value of the T^2_A will be performed. Set to 1 by default
gspe	A number indicating the term that will tune the spacing between steps for the SPE. Set to 1 by default (linear spacing).
gt2	A number indicating the term that will tune the spacing between steps for the SPE. Set to 1 by default (linear spacing).
mode	A character indicating the type of shift that will be performed: "simple", "steps" or "grid".

Value

list with elements: X, matrix with the new and shifted data, SPE and T2 vectors with the statistic values of each one of the new generated outliers or observations, elements `step.spe` and `step.t2` make reference to the step of each observation. Finally, the element `tag`, is a vector of ones as long as the number of generated observations.

scoutgrid

Shift of an array following a grid pattern.

Description

Shift of an array following a grid pattern.

Usage

```
scoutgrid(
  X,
  pcaref,
  T2.target = NA,
  SPE.target = NA,
  nsteps.t2 = 1,
  nsteps.spe = 1,
  gspe = 1,
  gt2 = 1
)
```

Arguments

X	Matrix with observations that will be shifted as rows.
pcaref	List with the elements of a PCA model: m (mean), s (standard deviation), prepro (preprocessing: "none", "cent" or "autosc"), P (loading matrix), lambda (vector with variances of each PC).
T2.target	A number indicating the target value for the T^2_A after the shift. Set to NA by default.
SPE.target	A number indicating the target value for the SPE after the shift. Set to NA by default.
nsteps.t2	An integer indicating the number of steps in which the shift from the reference to the target value of the T^2_A will be performed. Set to 1 by default.
nsteps.spe	An integer indicating the number of steps in which the shift from the reference to the target value of the SPE will be performed. Set to 1 by default.
gspe	A number indicating the term that will tune the spacing between steps for the SPE. Set to 1 by default (linear spacing).
gt2	A number indicating the term that will tune the spacing between steps for the SPE. Set to 1 by default (linear spacing).

Value

list with elements: `X`, matrix with the new and shifted data, `SPE` and `T2` vectors with the statistic values of each one of the new generated outliers or observations, elements `step.spe` and `step.t2` make reference to the step of each observation. Finally, the element `tag`, is a vector of ones as long as the number of generated observations.

<code>scoutsimple</code>	<i>Shift of an array with a single step.</i>
--------------------------	--

Description

Shift of an array with a single step.

Usage

```
scoutsimple(X, pcaref, T2.target = NA, SPE.target = NA)
```

Arguments

<code>X</code>	Matrix with observations that will be shifted as rows.
<code>pcaref</code>	List with the elements of a PCA model: <code>m</code> (mean), <code>s</code> (standard deviation), <code>prepro</code> (preprocessing: "none", "cent" or "autosc"), <code>P</code> (loading matrix), <code>lambda</code> (vector with variances of each PC).
<code>T2.target</code>	A number indicating the target value for the T^2_A after the shift. Set to NA by default.
<code>SPE.target</code>	A number indicating the target value for the SPE after the shift. Set to NA by default.

Value

list with elements: `X`, matrix with the new and shifted data, `SPE` and `T2` vectors with the statistic values of each one of the new generated outliers or observations, elements `step.spe` and `step.t2` make reference to the step of each observation. Finally, the element `tag`, is a vector of ones as long as the number of generated observations.

<code>scoutsteps</code>	<i>Shift of an array following a step-wise pattern.</i>
-------------------------	---

Description

Shift of an array following a step-wise pattern.

Usage

```
scoutsteps(
  X,
  pcaref,
  T2.target = NA,
  SPE.target = NA,
  nsteps = 1,
  gspe = 1,
  gt2 = 1
)
```

Arguments

X	Matrix with observations that will be shifted as rows.
pcaref	List with the elements of a PCA model: m (mean), s (standard deviation), prepro (preprocessing: "none", "cent" or "autosc"), P (loading matrix), lambda (vector with variances of each PC).
T2.target	A number indicating the target value for the T^2_A after the shift. Set to NA by default.
SPE.target	A number indicating the target value for the SPE after the shift. Set to NA by default.
nsteps	An integer indicating the number of steps in which the shift from the reference to the target values of the SPE and the T^2_A will be performed. Set to 1 by default.
gspe	A number indicating the term that will tune the spacing between steps for the SPE. Set to 1 by default (linear spacing).
gt2	A number indicating the term that will tune the spacing between steps for the SPE. Set to 1 by default (linear spacing).

Value

list with elements: X, matrix with the new and shifted data, SPE and T2 vectors with the statistic values of each one of the new generated outliers or observations, elements step.spe and step.t2 make reference to the step of each observation. Finally, the element tag, is a vector of ones as long as the number of generated observations.

speinfo	<i>Information about the Squared Prediction Error (SPE) of an observation. Two subplots show the information of an observation regarding its SPE statistic, i.e.: a bar plot indicating the value of the statistic for the observation, and a bar plot with the contribution that each component had for the SPE value</i>
---------	--

Description

Information about the Squared Prediction Error (SPE) of an observation. Two subplots show the information of an observation regarding its SPE statistic, i.e.: a bar plot indicating the value of the statistic for the observation, and a bar plot with the contribution that each component had for the SPE value

Usage

```
speinfo(SPE, E, limspe, iobs = NA)
```

Arguments

SPE	Vector with values of the SPE statistic.
E	Matrix with the contributions of each variable (columns) for each observation (rows) to the SPE. It is the error term obtained from the unexplained part of X by the PCA model.
limspe	Upper Control Limit for the SPE, at a certain confidence level $(1-\alpha)*100\%$.
iobs	Integer with the index of the observation of interest. Default value set to NA.

Value

ggplot object with the generated bar plots.

xshift	<i>Shift of an observation. The performed operation results as a combination of two main directions: the direction of maximum gradient for the SPE (weighted by the parameter b) and the direction of the projection of the observation on the model (weighted by the parameter a).</i>
--------	---

Description

Shift of an observation. The performed operation results as a combination of two main directions: the direction of maximum gradient for the SPE (weighted by the parameter b) and the direction of the projection of the observation on the model (weighted by the parameter a).

Usage

```
xshift(X, P, a, b)
```

Arguments

X	Matrix with observations that will be shifted
P	Loading matrix of the PCA model according to which the shift will be performed.
a	A number or vector tuning the shift in the direction of its projection.
b	A number or vector tuning the shift in the direction of its residual.

Value

Matrix with shifted observation as rows, keeping the order of the input matrix X.

Index

barwithucl, [2](#)
custombar, [2](#)
distplot, [3](#)
distplotsimple, [4](#)
ht2info, [5](#)
obscontribpanel, [5](#)
pcamb_classic, [6](#)
pcame, [6](#)
scoreplot, [7](#)
scoreplotsimple, [8](#)
scout, [9](#)
scoutgrid, [10](#)
scoutsimple, [11](#)
scoutsteps, [11](#)
speinfo, [12](#)
xshift, [13](#)