

# Package ‘SDisc’

February 14, 2012

**Type** Package

**Title** Integrated methodology for the identification of homogeneous subtypes in data

**Version** 1.24

**Date** 2011-10-27

**Author** Fabrice Colas

**Maintainer** Fabrice Colas <[sdisc@grano-salis.fr](mailto:sdisc@grano-salis.fr)>

**Description** Tools and methods to identify homogeneous subtypes in data by cluster analysis; includes methods for data pre-processing, repeated cluster analysis, model selection, model reliability and reproducibility assessment, subtype characterization and validation.

**Depends** R (>= 2.12.1), mclust, stats, utils, RColorBrewer, abind, xtable, digest, e1071, snow, SparseM

**License** BSD

**URL** <http://grano-salis.fr/sdisc/>

**Repository** CRAN

**Date/Publication** 2011-11-02 17:17:14

## R topics documented:

agreementScores . . . . .	2
bestModel . . . . .	2
bicTable . . . . .	3
expDesign . . . . .	4
modelBasedEM . . . . .	5
naPattern . . . . .	6
paste2 . . . . .	7
SDCModel . . . . .	7
SDData . . . . .	9
SDDataSettings . . . . .	11

SDisC	11
SDisCReportHead	14
SDisCStability	15
texTable	16
write.SDisC	16

<b>Index</b>	<b>18</b>
--------------	-----------

---

agreementScores	<i>Agreement scores between two classifiers</i>
-----------------	---

---

### Description

Given two classification results, this function returns different statistics to assess the level of agreement: the random index, kappa, V score, the chi2 p-value and the grand total.

### Usage

```
agreementScores(x, y=NULL)
```

### Arguments

x	classification of the first classifier
y	classification of the second classifier

### Author(s)

Fabrice Colas

### See Also

[predict.SDisC](#), [print.SDisC](#), [summary.SDisC](#),

---

bestModel	<i>Most likely mixture models of an SDisC analysis</i>
-----------	--

---

### Description

Taking as input an SDisC analysis, returns a character vector of the most likely models ordered by decreasing likelihood of length n. Optional parameters modelName and G may act as filter to subset the most likely model given those parameters.

### Usage

```
bestModel(x, n=NULL, modelName=NULL, G=NULL)
```

**Arguments**

x	an SDisc object
n	the number of model to return, by default 5 as specified in SDisc
modelName	filter the returned models matching modelName
G	filter the returned models to those matching with G

**Author(s)**

Fabrice Colas

**See Also**

[summary.SDCModel](#),

---

bicTable	<i>BIC table of an SDisc analysis</i>
----------	---------------------------------------

---

**Description**

Returns for each combination of model, number of mixtures and random initialization seed the BIC score (Bayesian Information Criterion). Besides the absolute BIC score, we report, too, those same BIC score but relatively to absolute BIC score of the most likely model.

**Usage**

```
## Default S3 method:
bicTable(x, ...)
## S3 method for class 'bicTable'
print(x, n=NULL, modelName=NULL, G=NULL, latex=FALSE, lab="bic5", ...)
## S3 method for class 'bicTable'
summary(object, fun="min", bic="relativeBic", latex=FALSE, lab="bic", fmt="%.2f", ...)
```

**Arguments**

x	an SDisc analysis
n	either null for the whole BIC table, or a number between 0 and 1 for those models having a relative difference at most n% worst than the most likely one (BIC), or a number greater than one returning the ordered table of the best n models.
modelName	limit the table to those models of type modelName
G	limit the table to those models with G components
latex	whether the table must be returned as a LaTeX source code (for dynamic reporting)
lab	the label of the latex table

object	a bicTable object
fun	the name of the function call for the main estimate of the table, by default the min relative BIC is searched for
bic	either <code>relativeBic</code> -by default-, or BIC
fmt	the format for the numbers in the table, by default two digits after the comma (see <code>sprintf</code> )
...	additional parameters passed to subfunctions, e.g. <code>texTable</code>

### Details

`summary` returns an aggregated summary of the repeated data modeling performed for different combinations of number of mixture components, number of mixture model parameters and random initialization start. For each combination, it reports the best BIC score relatively to the top-ranking one along with the 95% range of BIC scores.

### Author(s)

Fabrice Colas

### See Also

[bicTable](#), [SDisc](#), [texTable](#)

### Examples

```
settings <- SDDataSettings(iris)
settings['Species',] <- c(NA,FALSE, NA, NA, NA, NA)
x <- SDisc(iris, settings=settings, prefix='iris')
bicTable(x)
```

---

expDesign

*Charts the Joint Distribution of Groups of Factors Two by Two*

---

### Description

The method enables to report the joint distributions of two or more factors with each others. If `gr1` and `gr2` are both `NULL`, then each column of `x` is compared two-by-two. If `gr1` is not `NULL` but `gr2` is, then the different factors of `gr1` are compared one to another. Finally, if `gr2` is not null, the factors of `gr1` are compared with those of `gr2`.

### Usage

```
expDesign(x, gr1 = NULL, gr2 = NULL, tex = FALSE, exclude = NULL)
```

**Arguments**

x	a data frame or matrix,
gr1	a set of factor names from x, NULL by default
gr2	a second set of factor names from x to compare against the factors from gr1, NULL by default
tex	if TRUE a (Sweave-friendly) latex output is generated; if FALSE tables are directly returned to the standard output
exclude	should missing values be considered as level; by default yes NULL -see help of table-

**Author(s)**

Fabrice Colas

**Examples**

```
expDesign(OrchardSprays, gr1='treatment', gr2='rowpos')
expDesign(OrchardSprays, gr1='treatment', gr2='rowpos', tex=TRUE)
```

---

 modelBasedEM

---

*Model based clustering function starting with an E-step*


---

**Description**

Proceeds to an analysis for cluster by mixture modeling (mclust). It takes as input a SDCModel (x) and a data matrix data. To init EM, a cluster membership probability matrix (z) is drawn at random (NxG with N the number of observations and G the number of clusters). The corresponding model is estimated by an M-step and then, EM is started.

**Usage**

```
modelBasedEM(x, data)
```

**Arguments**

x	an SDCModel, e.g. the first element of an SDisc()[[1]] analysis,
data	a data matrix, e.g. the one provided by print(SDData(x))

**Author(s)**

Fabrice Colas

## References

Fraley C, Raftery AE: Model-Based Clustering, Discriminant Analysis and Density Estimation. Journal of the American Statistical Association, vol. 97, pp. 611-631, 2002. [<http://www.stat.washington.edu/raftery/Research/PDF/fraley2002.pdf>]

Fraley C, Raftery AE, MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering. Technical Report 504, Department of Statistics, University of Washington, September 2006. [<http://www.stat.washington.edu/fraley/mclust/tr504.pdf>]

## See Also

[print.SDCModel](#), [print.SDData](#), [SDisc](#), [SDStability](#),

---

naPattern

*Pattern of missing value of a data set*

---

## Description

It either returns a character vector of the record ID's presenting missing values or a table (`latex=TRUE`) report of the missingness pattern for each record presenting at least one missing value .

## Usage

```
naPattern(x, latex=FALSE, ...)
```

## Arguments

<code>x</code>	an SDData object
<code>latex</code>	by default, FALSE
<code>...</code>	additional arguments to be passed to <code>texTable</code> .

## Author(s)

Fabrice Colas

## See Also

[SDData](#), [texTable](#),

## Examples

```
settings <- SDDataSettings(iris)
settings['Species',] <- c(NA,FALSE, NA, NA, NA, NA)
x <- SDData(iris, settings=settings, prefix='iris')
naPattern(x)
```

---

paste2	<i>Concatenate strings without any space</i>
--------	--

---

**Description**

paste2 is a wrapper to paste that recursively concatenate all the string without any space nor separators.

**Usage**

```
paste2(...)
```

**Arguments**

... one or more R objects, to be converted to character vectors.

**Author(s)**

Fabrice Colas

**See Also**

[paste](#)

**Examples**

```
paste2(seq(1:9))
```

---

SDCModel	<i>SDCModel, SDisc's Cluster Model class</i>
----------	--

---

**Description**

SDCModel is the name given to SDisc Cluster Models. Subtype models from an SDisc analysis can be accessed as elements of a list, via their name or an index integer. With print, we return the mixture component (subtype) membership likelihood per sample as well as the mapping for the most likely mixture. The summary method characterizes the model in terms of odd ratio and joint distribution statistics with respect to, for instance, an confounding or verifactory factor -like gender or strata- that was not included in the subtype modeling.

**Usage**

```
## S3 method for class 'SDCModel'  
print(x, data, ...)  
## S3 method for class 'SDCModel'  
summary(object, data, type="oddRatiosB", latex=FALSE, lab="", shortStr=FALSE, ...)
```

**Arguments**

<code>x</code>	an SDCModel
<code>data</code>	an SDData container
<code>object</code>	an SDCModel obtained from (see <code>SDisc</code> and <code>bestModel</code> )
<code>type</code>	the type of summary to report, in: "oddRatios", "oddRatiosB" by default, "chi2test", "jointDistrib"
<code>lab</code>	the label of the LaTeX table
<code>shortStr</code>	whether to shorten the names of the different factors by their first letters
<code>latex</code>	whether a LaTeX formatted output should be generated
<code>...</code>	(currently not used) additional parameters passed to the sub-functions

**Details**

`print` returns a matrix with  $G+1$  columns. The  $G$  first columns represent the likelihood of each record to belong to the different mixture components. The last column corresponds the mapping of each record into one of the  $G$  components.

`summary` returns different statistical summaries on a SDCModel. Among the possible summaries there are odd ratio, chi2 statistics and joint distribution with a categorical target variable. Odd ratios statistics are computed on factors of variable defined in the settings configuration file (see `oddGroup` in `SDDataSettings`). Two types of odd ratios are implemented: the `oddRatios` cross-product is based on counts of the number of occurrence above and below the mean of the data distribution, while for `oddRatiosB` the effect size is used in the cross-product. Concerning `chi2test` and `jointDistrib`, they involve a comparison with a target categorical variable.

**Author(s)**

Fabrice Colas

**See Also**

[modelBasedEM](#), [SDDataSettings](#), [texTable](#)

**Examples**

```
settings <- SDDataSettings(iris)
settings['Species',] <- c(NA,FALSE, NA, NA, NA, NA)
x <- SDisc(iris, settings=settings, prefix='iris')
xBestModel <- x[[bestModel(x,1)]]
print(xBestModel, data=SDData(x))
summary(x[[bestModel(x,1)]], data=SDData(x))
```

---

SDData *Data container for SDisc analyses*

---

### Description

SDisc dataset container constructor taking as input a description of the data and an analysis prefix. SDData and its methods transform, print, summary and plot the data.

### Usage

```
## Default S3 method:
SDData(x, prefix, dataOrig=NULL, TData=NULL, settings=NULL, initFun=list(SDDataCC), subset=NULL, ...)
## S3 method for class 'SDData'
print(x, rseed=NULL, range=1:3, allNumVars = FALSE, latex=FALSE, ...)
## S3 method for class 'SDData'
plot(x, q=NULL, est = 1, zlim = c(-2, 2), latex=FALSE, ...)
## S3 method for class 'SDData'
summary(object, q=NULL, latex=FALSE, digits = 3, ...)
## S3 method for class 'SDData'
predict(object, newdata, prefix = "Newdata", subset = NULL, ...)
```

### Arguments

x	a data matrix, a previously instantiated SDData container or an SDisc object from which the SDData will be extracted,
dataOrig	original data, NULL by default
TData	a set of operations TData to apply to the data
settings	a data matrix as generated by SDDataSettings or the path to a CSV file separated by ";"
initFun	a list of function taking a data matrix as input and returning a treated matrix as output. As default, returns all complete cases records
prefix	a prefix that will serve to identify the analysis in the dynamic report and to define the storage place
subset	a subset of record indeces for the data set (row names)
rseed	an integer to set the random number generator that will select randomly a set of rows and columns to see before and after the data treatment, the data matrix
range	a sequence of integers to subset the randomly ordered column and row names vectors
allNumVars	whether all numeric variables should be printed, inclusive those not necessarily included in the cluster modeling; these variables are retrieved from dataOrig
q	limit the summary to a subset of the data treatments (regular expression). When a character vector is provided, the data treatments matching the regular expression are plotted side by side in an image.
est	TODO

<code>zlim</code>	limits for the heatmap
<code>object</code>	an data set container SDData
<code>digits</code>	how many digits should be reported in the SDData summary
<code>newdata</code>	a new dataset on which to apply the same transformation estimated on the first SDData
<code>latex</code>	whether the table must be returned into a LaTeX code
<code>...</code>	additional parameters to be passed to the subfunctions

### Details

SDData is the data container constructor for SDisc analyses. It proceeds copying the original data, creating the working directories for figures and tables, and archive the dataset as an RData file. Default dataset initialization function (`initFun`) filter out incomplete cases because clustering algorithm used in SDisc (`McLust`) does not handle missing values. To select a data subset from a previous SDData, a selection index can also be passed.

`predict` transforms a new data set based on the transformation estimates -like mean, standard deviation- from another SDData container.

`print` returns the data matrix that results from the initialization (e.g. complete cases) and the data treatments applied on the different variables, as defined in the data settings configuration file. To verify the data treatments, an `rseed` can be provided to place in parallel a random extract of the data matrix before and after the data treatments. The `range` parameter gives the number of rows and columns to extract randomly.

`plot` reports as PDF the boxplots and histograms for each variable of the data set container. If `latex` is set to TRUE, then LaTeX code for Sweave vignettes is returned.

`summary` returns a summary of the data treatments operated on the data set. For `mean`, `sd`, `scale`, etc, it returns the estimates. For `lm`, it returns the estimates of the coefficients along with their standard error,  $p$ -value, the  $R^2$  and adjusted  $R^2$  of the transformation, and the number of records on which the estimate was based on.

### Author(s)

Fabrice Colas

### See Also

[modelBasedEM](#), [SDDataSettings](#), [naPattern](#), [SDDataSettings](#), [SDisc](#), [texTable](#), [SDData](#)

### Examples

```
settings <- SDDataSettings(iris)
settings['Species',] <- c(NA,FALSE, NA, NA, NA,NA)
x <- SDData(iris, settings=settings, prefix='iris')
summary(x)
### DO NOT RUN
# plot(x)
```

**Description**

Set or returns the data settings that are used to define the data: how they must be represented graphically (heatmap, parallel coordinates), what are the meaningful factors of variables, what are the data treatments.

If a matrix is provided, it generates a 'working' settings file (settings.csv) for the current data matrix. It can be saved asCSV or returned to the standard output.

In the case an SDisc or an SDData is provided, it returns the settings definition of the current data container or SDisc analysis.

**Usage**

```
SDDataSettings(x, asCSV=FALSE, inCAalysis=NULL, latex=FALSE)
```

**Arguments**

x	a matrix, an SDData or an SDisc object
asCSV	whether to save the settings as a CSV file
inCAalysis	a character vector to specify the set of variables to include in the mixture modeling
latex	whether a LaTeX output should be generated

**Author(s)**

Fabrice Colas

**See Also**

[plot.SDData](#), [plot.SDisc](#), [print.SDData](#), [SDData](#), [SDisc](#), [summary.SDData](#), [texTable](#),

**Description**

Performs and returns an SDisc analysis on the data. This analysis involves repeated mixture modeling for different combinations of number of components, number of mixture model parameters and random initialization start. The SDisc results contains a data set container (SDData) which stores the original data and which may, e.g., limit the cluster analysis to a few variables (see SDData). It contains, too, the parameters of the different models estimated and a BIC table summarizing their likelihood and rank. Generic plot, print and summary function enable to visualize and summarize the results (see [plot.SDisc](#), [print.SDisc](#) and [summary.SDisc](#)).

**Usage**

```
## Default S3 method:
SDisc(x, cfun="modelBasedEM", cFunSettings=list(modelName = c("EII", "VII"), G = 3:5, rseed = 6013:6015)

## S3 method for class 'SDisc'
plot(x, q=NULL, type=c('plotParcoord', 'plotLegend', 'plotPC1', 'plotPC2', 'plotDendroCluster',
  'plotDendroVar'), latex=FALSE, title=NULL, xlim=c(-3, 3), zlim=c(-2, 2), xy=c(-2.2, 0), pattern=mean,
  colGrad=rev(brewer.pal(9, "RdBu")), rangeFV=NULL, lab=NULL, ...)
## S3 method for class 'SDisc'
predict(object, newdata, ...)
## S3 method for class 'SDisc'
print(x, y=NULL, m1=1, m2=2, latex=FALSE, lab="jointdistrib", ...)
## S3 method for class 'SDisc'
summary(object, q = 1, ...)
```

**Arguments**

x	a data matrix (with its settings file), an SDData instance or an SDisc data object
cfun	the name of the cluster algorithm
cFunSettings	the set of parameters of the cluster algorithm
nTopModels	the number of top-ranking models
nnodes	the number of nodes in the case of parallel computing
object	an SDisc analysis result
q	an numeric value telling how many top ranking models to characterize graphically (bestModel, a character vector referring to the names of an SDCModel, by default set to NULL that is, the 5 most likely models)
y	an optional second SDisc data object whose model where estimated on the same data
m1	rank (integer) or name (character vector) of the first model to compare. The rank is passed to bestModel to retrieve the appropriate model name.
m2	rank (integer) or name (character vector) of the second model to compare. The rank is passed to bestModel to retrieve the appropriate model name.
latex	either TRUE or FALSE, whether the LaTeX code must be reported on the standard output for dynamic report generation (Sweave)
lab	the label of that table
type	a character vector in 'plotParcoord', 'plotLegend', 'plotImage', 'plotDendroCluster', 'plotDendroVar'
title	the title of the graphics
xlim	the x-limits of the parallel coordinate plots
zlim	the z-limits for the color gradient in the image
xy	the xy-location of the legend

pattern	the name of the function to calculate the characteristic pattern, by default mean
cex	the character expansion numeric value, by default 0.7 70%
colGrad	a character vector of the colors to use in the color image
rangeFV	the range of features when plotting series of values
newdata	an SDData or SDisc object
...	additional parameters passed to the SDData function when a data matrix is provided. Such parameters may be settings and prefix.

### Details

plot characterizes graphically the best model ( $q=NULL$ ) of an SDisc analysis through parallel coordinate, legend, image, sample- and variable-dendrograms. Some plots like heatmap and parallel coordinates admit parameterization, which can be specified in the data description (see `SDDataSettings(x)`). Generated graphics are in PDF and Sweave output can be switched on with `latex=TRUE`. Additional parameters influence the color scales, range of values.

predict estimates a set of subtypes (mixture models) on new data based on the most likely mixture model from an SDisc analysis. Used in combination with print, the new data can help validate a previous model of subtypes.

print reports the joint distribution between pairs of subtyping results, the agreement ( $\kappa$ , V  $\chi^2$ -based measure, random index). It is also possible to compare affectation between SDisc analyses with different parameters but with the same data container. In which case, it illustrates how, under different settings, the subtyping analyses agree.

summary lists the dataset characteristics, the top-ranked models, the BIC table, and oddratios statistics for the top ranked models.

### Author(s)

Fabrice Colas

### References

Colas, F, Meulenbelt, I, Houwing-Duistermaat, JJ, Kloppenburg, M, Watt, I, van Rooden, SM, Visser, M, Marinus, H, Cannon, EO, Bender, A, van Hilten, JJ, Slagboom, PE, Kok, JN: A Scenario Implementation in R for Subtype Discovery Exemplified on Chemoinformatics Data. Leveraging Applications of Formal Methods, Verification and Validation (ISoLA'08), October 13-15, 2008. [[http://dx.doi.org/10.1007/978-3-540-88479-8\\_48](http://dx.doi.org/10.1007/978-3-540-88479-8_48)]

Colas, F, Meulenbelt, I, Houwing-Duistermaat, JJ, Kloppenburg, M, Watt, I, van Rooden, SM, Visser, M, Marinus, H, van Hilten, JJ, Slagboom, PE, Kok, JN: Stability of Clusters for Different Time Adjustments in Complex Disease. Research 30th Annual International IEEE EMBS Conference (EMBC'08), Vancouver, Canada, 2008. [<http://dx.doi.org/10.1109/IEMBS.2008.4650238>]

Fraley C, Raftery AE: Model-Based Clustering, Discriminant Analysis and Density Estimation. Journal of the American Statistical Association, vol. 97, pp. 611-631, 2002. [<http://www.stat.washington.edu/raftery/Research/PDF/fraley2002.pdf>]

Fraley C, Raftery AE, MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering. Technical Report 504, Department of Statistics, University of Washington, September 2006. [<http://www.stat.washington.edu/fraley/mclust/tr504.pdf>]

**See Also**

[SDStability](#), [SDDData](#), [SDDDataSettings](#), [SDisc](#), [bestModel](#), [bicTable](#), [texTable](#), [agreementScores](#)

**Examples**

```
settings <- SDDDataSettings(iris)
settings['Species',] <- c(NA,FALSE, NA, NA, NA, NA)
x <- SDisc(iris, settings=settings, prefix='iris')
### do not run
#plot(x)
#print(x)
summary(x)
```

---

SDiscReportHead

*SDiscReportHead*


---

**Description**

LaTeX header for Sweave reporting.

**Usage**

```
SDiscReportHead(LO='Report', packages=list('Sweave', 'amsmath', 'underscore', 'setspace', 'pdfscape',
'multirow', 'glossaries'=c('toc', 'acronym', 'xindy'), 'multicol', inputenc='latin1', babel='englis
'pdfpages', caption=c('small', 'bf'), 'graphicx', 'fancyhdr', 'lastpage', 'longtable', 'color', xcolor=
geometry=c('left=1.25cm', 'top=2cm', 'right=1.25cm', 'bottom=2cm'), hyperref=c('colorlinks=true',
'citecolor=blueDoc', 'filecolor=blueDoc', 'linkcolor=blueDoc',
'urlcolor=blueDoc'), 'lscape', 'sectsty', 'colortbl', 'wrapfig', 'array'),
author=list(name='MyName', email='MyEmail', address='address'))
```

**Arguments**

LO	the top left page header (default is "Report")
packages	named list of packages. List names refer to LaTeX packages, and are passed to the usepackage command. For each name, the character vector defines the list of options (squared brackets) of the LaTeX package, comma separated.
author	defines the author list; for one author, a list of three elements (name, email, address); for more authors, a list element of lists composed of (name, email, address).

**Author(s)**

Fabrice Colas

---

*SDStability**SDStability*

---

**Description**

Evaluates the stability of a subtype model by introducing step by step more noise in the data.

**Usage**

```
## Default S3 method:  
SDStability(x, q, rseed=6013, nnoise=10, nrep=10)  
## S3 method for class 'SDStability'  
plot(x, ncolors=9, minmax=c(0.5, 1), ...)
```

**Arguments**

x	an SDisc analysis result
q	the model to assess
rseed	an integer to initialize the random number generator
nnoise	an integer to set the number of noise levels
nrep	an integer to indicate the number of replicates
ncolors	the number of colors in the grey gradient
minmax	the range of values for the color, by default 0.5 to 1 of the $V$ values.
...	(currently not use) additional parameters passed to the graphic functions

**Details**

plot illustrates by color images the result of stability analyzes (SDStability).

**Author(s)**

Fabrice Colas

**References**

Colas, F, Meulenbelt, I, Houwing-Duistermaat, JJ, Kloppenburg, M, Watt, I, van Rooden, SM, Visser, M, Marinus, H, van Hilten, JJ, Slagboom, PE, Kok, JN: Stability of Clusters for Different Time Adjustments in Complex Disease. Research 30th Annual International IEEE EMBS Conference (EMBC'08), Vancouver, Canada, 2008. [<http://dx.doi.org/10.1109/IEMBS.2008.4650238>]

---

texTable	<i>texTable</i>
----------	-----------------

---

**Description**

texTable

**Usage**

```
texTable(x, cap="", lab="", digits=NULL, sanitize=TRUE, align = NULL, oddColor = "blueLines", longtab =
```

**Arguments**

x	a matrix
cap	the caption of the LaTeX table
lab	the label of the LaTeX table
digits	the number of digits to display
sanitize	whether to sanitize the row and column names values for LaTeX table (xtable). If so, define an alternative treatment function,
align	the alignment vector of the table
oddColor	the selected color to highlight alternatively the rows of the tables
longtab	whether a LaTeX long table environment should be used for this table
...	additional parameters to be passed to xtable

**Author(s)**

Fabrice Colas

**See Also**

[print.bicTable](#), [print.SDData](#), [print.SDisc](#), [SDDataSettings](#), [summary.bicTable](#), [summary.SDCModel](#), [summary.SDData](#), [summary.SDisc](#),

---

write.SDisc	<i>write.SDisc</i>
-------------	--------------------

---

**Description**

For the 5 best models (by default, see nTopModel in SDisc), writes as CSV files the cluster membership matrix of the different records along with their affectation into the different clusters (see print.SDCModel). By default, SDisc performs this task and save the models.

**Usage**

```
write.SDisc(x, q=NULL)
```

**Arguments**

x	an SDisc analysis result
q	a NULL, numeric or character vector as provided by bestModel.

**Author(s)**

Fabrice Colas

**See Also**

[print.SDisc](#), [SDisc](#),

**Examples**

```
settings <- SDDataSettings(iris)
settings['Species',] <- c(NA,FALSE, NA, NA, NA, NA)
x <- SDisc(iris, settings=settings, prefix='iris')
write.SDisc(x)
```

# Index

- \*Topic **NA**
    - naPattern, 6
  - \*Topic **arith**
    - bestModel, 2
  - \*Topic **attribute**
    - SDDataSettings, 11
  - \*Topic **character**
    - bestModel, 2
  - \*Topic **classif**
    - SDisc, 11
  - \*Topic **cluster**
    - bicTable, 3
    - modelBasedEM, 5
    - SDCModel, 7
    - SDisc, 11
    - write.SDisc, 16
  - \*Topic **data**
    - SDData, 9
  - \*Topic **distribution**
    - agreementScores, 2
  - \*Topic **hplot**
    - SDData, 9
    - SDisc, 11
    - SDStability, 15
  - \*Topic **htest**
    - SDCModel, 7
    - SDData, 9
  - \*Topic **manip**
    - SDData, 9
    - SDDataSettings, 11
    - texTable, 16
  - \*Topic **math**
    - SDData, 9
    - SDisc, 11
  - \*Topic **methods**
    - bicTable, 3
    - SDCModel, 7
    - SDData, 9
    - SDDataSettings, 11
    - SDisc, 11
    - SDStability, 15
  - \*Topic **models**
    - modelBasedEM, 5
    - SDData, 9
    - SDisc, 11
  - \*Topic **multivariate**
    - modelBasedEM, 5
    - SDData, 9
    - SDisc, 11
  - \*Topic **print**
    - bicTable, 3
    - SDCModel, 7
    - SDData, 9
    - SDisc, 11
    - SDiscReportHead, 14
    - texTable, 16
    - write.SDisc, 16
  - \*Topic **regression**
    - SDData, 9
  - \*Topic **robust**
    - SDStability, 15
  - \*Topic **utilities**
    - bicTable, 3
    - write.SDisc, 16
- agreementScores, 2, 14
- bestModel, 2, 14
- bicTable, 3, 4, 14
- compareModel (SDisc), 11
- expDesign, 4
- jointDistrib (SDCModel), 7
- modelBasedEM, 5, 8, 10
- naPattern, 6, 10

oddRatios (SDCModel), 7  
oddRatiosB (SDCModel), 7

paste, 7  
paste2, 7  
plot.SDData, 11  
plot.SDData (SDData), 9  
plot.SDisc, 11  
plot.SDisc (SDisc), 11  
plot.SDStability (SDStability), 15  
plotDendroCluster (SDisc), 11  
plotDendroVar (SDisc), 11  
plotImage (SDisc), 11  
plotLegend (SDisc), 11  
plotParcoord (SDisc), 11  
plotSeries (SDisc), 11  
predict.SDData (SDData), 9  
predict.SDisc, 2  
predict.SDisc (SDisc), 11  
print.bicTable, 16  
print.bicTable (bicTable), 3  
print.SDCModel, 6  
print.SDCModel (SDCModel), 7  
print.SDData, 6, 11, 16  
print.SDData (SDData), 9  
print.SDisc, 2, 16, 17  
print.SDisc (SDisc), 11

SDCModel, 7  
SDData, 6, 9, 10, 11, 14  
SDDataSettings, 8, 10, 11, 14, 16  
SDisc, 4, 6, 10, 11, 11, 14, 17  
SDiscReportHead, 14  
SDStability, 6, 14, 15  
summary.bicTable, 16  
summary.bicTable (bicTable), 3  
summary.SDCModel, 3, 16  
summary.SDCModel (SDCModel), 7  
summary.SDData, 11, 16  
summary.SDData (SDData), 9  
summary.SDisc, 2, 16  
summary.SDisc (SDisc), 11

texTable, 4, 6, 8, 10, 11, 14, 16

write.SDisc, 16