Package 'RSDA'

July 2, 2014

Type Package

Title RSDA - R to Symbolic Data Analysis

Version 1.2

Date 2014-03-18

Author Oldemar Rodriguez R. with contributions from Olger Calderon and Roberto Zuniga

Maintainer Oldemar Rodriguez <oldemar.rodriguez@ucr.ac.cr>

Description Symbolic Data Analysis (SDA) was proposed by professor Edwin Diday in 1987, the main purpose of SDA is to substitute the set of rows (cases) in the data table for a concept (second order statistical unit). This package implements, to the symbolic case, certain techniques of automatic classification, as well as some linear models.

License GPL (>= 2)

Depends R (>= 2.10), sqldf, XML, scales, FactoMineR, ggplot2

Imports glmnet, abind, scatterplot3d

URL http://www.oldemarrodriguez.com

NeedsCompilation no

Repository CRAN

Date/Publication 2014-03-18 18:20:41

R topics documented:

`abalone` *SODAS XML data file.*

Description

Example of SODAS XML data file converted in a CSV file in RSDA format.

Usage

```
data(abalone)
```

Format

abalone<-SODAS.to.RSDA("C:/Program Files (x86)/DECISIA/SODAS version 2.0/bases/abalone.xml)

Source

http://www.info.fundp.ac.be/asso/sodaslink.htm

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Examples

```
# We can read the file directly from the SODAS XML file as follows:
# abalone<-SODAS.to.RSDA("C:/Program Files (x86)/DECISIA/SODAS version 2.0/bases/abalone.xml)
# We can save the file in CSV to RSDA format as follows:
# write.sym.table(sodas.ex1, file='abalone.csv', sep=';',dec='.', row.names=TRUE,
#               col.names=TRUE)
# We read the file from the CSV file,
# this is not necessary if the file is read directly from
# XML using SODAS.to.RSDA as in the first statement in this example.
data(abalone)
res<-sym.interval.pca(abalone,'centers')
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
               labels=TRUE,col='red',main='PCA Oils Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                 sym.var(res$Sym.Components,3),color='blue',main='PCA Oils Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                      labels=TRUE)
sym.circle.plot(res$Sym.Prin.Correlations)
```

An introduction to RSDA

R to Symbolic Data Analysis

Description

This work is framed inside the Symbolic Data Analysis (SDA). The objective of this work is to implement in R to the symbolic case certain techniques of the automatic classification, as well as some lineal models. These implementations will always be made following two fundamental principles in Symbolic Data Analysis like they are: Classic Data Analysis should always be a case particular case of the Symbolic Data Analysis and both, the exit as the input in an Symbolic Data Analysis should be symbolic. We implement for variables of type interval the mean, the median, the mean of the extreme values, the standard deviation, the deviation quartil, the dispersion boxes and the correlation also three new methods are also presented to carry out the lineal regression for variables of type interval. We also implement in this R package the method of Principal Components Analysis in two senses: First, we propose three ways to project the interval variables in the circle of correlations in such way that is reflected the variation or the inexactness of the variables. Second, we propose an algorithm to make the Principal Components Analysis for variables of type histogram. We implement a method for multidimensional scaling of interval data, denominated INTERSCAL.

Details

Package:	RSDA
Type:	Package
Version:	1.2
Date:	2014-03-18
License:	GPL (>=2)

Most of the function of the package stars from a symbolic data table that can be store in a CSV file withe follwing forma: In the first row the labels $C means that follows a continuous variable, $I means an interval variable, $H means a histogram variables and $S means set variable. In the first row each labels should be follow of a name to variable and to the case of histogram a set variables types the names of the modalities (categories) . In data rows for continuous variables we have just one value, for interval variables we have the minimum and the maximum of the interval, for histogram variables we have the number of modalities and then the probability of each modality and for set variables we have the cardinality of the set and next the elements of the set.

Author(s)

Oldemar Rodriguez Rojas

Maintainer: Oldemar Rodriguez Rojas <oldemar.rodriguez@ucr.ac.cr>

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Billard L., Douzal-Chouakria A. and Diday E. (2011) Symbolic Principal Components For Interval-Valued Observations, Statistical Analysis and Data Mining. 4 (2), 229-246. Wiley.

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Carvalho F., Souza R.,Chavent M., and Lechevallier Y. (2006) Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. Pattern Recognition Letters Volume 27, Issue 3, February 2006, Pages 167-179

Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales a des donnees de type intervalle, Rev. Statistique Appliquee, Vol. XLV Num. 3 pag. 5-24, France.

Diday, E., Rodriguez O. and Winberg S. (2000). Generalization of the Principal Components Analysis to Histogram Data, 4th European Conference on Principles and Practice of Knowledge Discovery in Data Bases, September 12-16, 2000, Lyon, France.

Chouakria A. (1998) Extension des methodes d'analysis factorialle a des donnees de type intervalle, Ph.D. Thesis, Paris IX Dauphine University.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

See Also

display.sym.table

Examples

```
data(example3)
sym.data<-example3
display.sym.table(sym.data)
sym.scatterplot(sym.var(sym.data,1),sym.var(sym.data,4),col='blue',main='Main Title')

data(oils)
res<-sym.interval.pca(oils,'centers')
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
  labels=TRUE,col='red',main='PCA Oils Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
   sym.var(res$Sym.Components,3),color='blue',main='PCA Oils Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),
   sym.var(res$Sym.Components,2), labels=TRUE)
sym.circle.plot(res$Sym.Prin.Correlations)

res<-sym.interval.pca(oils,'classic')
plot.PCA(res,choix="ind")
plot.PCA(res,choix="var")
```

```
data(lynne2)
res<-sym.interval.pca(lynne2,'centers')
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
  labels=TRUE, col='red',main='PCA Lynne Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
  sym.var(res$Sym.Components,3),color='blue', main='PCA Lynne Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
  labels=TRUE)
sym.circle.plot(res$Sym.Prin.Correlations)
```

`Cardiological` *Cardiological data example*

Description

Cardiological interval data example.

Usage

```
data(Cardiological)
```

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Examples

```
data(Cardiological)
car.data<-Cardiological
res.cm<-sym.lm(Pulse~Syst+Diast,sym.data=car.data,method='cm')
pred.cm<-predictsym.lm(res.cm,car.data,method='cm')
RMSE.L(sym.var(car.data,1),pred.cm$Fitted)
RMSE.U(sym.var(car.data,1),pred.cm$Fitted)
R2.L(sym.var(car.data,1),pred.cm$Fitted)
R2.U(sym.var(car.data,1),pred.cm$Fitted)
deter.coefficient(sym.var(car.data,1),pred.cm$Fitted)
```

`cfa.scatterplot` *CFA Symbolic Scatter Plot*

Description

This function could be use to plot two symbolic variables in a X-Y plane to special case of Symbolic Correspondance Analysis.

Usage

```
cfa.scatterplot(sym.var.x, sym.var.y, num.gr1=0, labels = TRUE, ...)
```

Arguments

`sym.var.x`	First symbolic variable
`sym.var.y`	Second symbolic variable.
`num.gr1`	Number of modes of the first variable
`labels`	As in R plot function.
`...`	As in R plot function.

Value

Return a graphics.

Author(s)

Oldemar Rodriguez Rojas

References

Rodriguez, O. (2011). Correspondence Analysis for Symbolic MultiValued Variables. Workshop in Symbolic Data Analysis Namur, Belgium.

See Also

`sym.cfa`

Examples

```
data(ex_cfa1)
res<-sym.cfa(ex_cfa1)
cfa.scatterplot(sym.var(res,1),sym.var(res,2),num.gr1=ex_cfa1$N,labels=TRUE,col='red',
                        main='CFA')
```

`classic.to.sym` *Generate a symbolic data table*

Description

Generate a symbolic data table from a classic data table.

Usage

```
classic.to.sym(dataTable, concept, variables, variables.types)
```

Arguments

`dataTable`	This is the classic data table.
`concept`	These are the variable that we are going to use a concepts.
`variables`	These are the variables that we want to inlude in the symbolic data table.
`variables.types`	These are the variables symbolic types (continuos, interval, set or histograma) of the variables that we want to inlude in the symbolic data table.

Value

The symbolic data table.

Author(s)

Olger Calderon and Roberto Zuniga.

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

`read.sym.table`

Examples

```
data(ex1_db2so)
ex1 <- ex1_db2so
result <- classic.to.sym(ex1, concept=c("state", "sex"),
                  variables=c("county", "group", "age","age"),
                  variables.types=c("$C", "$I", "$H", "$S"))
result
```

`deter.coefficient` *Compute the determination cosfficient*

Description

The determination coefficient represents a goodness-of-fit measure commonly used in regression analysis to capture the adjustment quality of a model.

Usage

```
deter.coefficient(sym.var, prediction)
```

Arguments

`sym.var`	Variable that was predicted.
`prediction`	The prediction given by the model.

Value

Return the determination cosfficient.

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

`sym.glm`

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
deter.coefficient(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
deter.coefficient(sym.var(int_prost_test,9),pred.cm.lasso)
```

`display.sym.table` *display.sym.table*

Description

This function display a symbolic data table tha have been read by read.sym.table(...)

Usage

```
display.sym.table(sym.data)
```

Arguments

`sym.data`	Shoud be a Symbolic Data table that have been read with read.sym.table(...)

Details

The output will be the symbolic data table in the screen.

Value

Not value.

Author(s)

Oldemar Rodriguez Rojas

References

Billard, L and Diday, E. (2007). Symbolic Data Analysis: Conceptual Statistics and Data Mining (Wiley Series in Computational Statistics).

Billard, L and Diday, E. (2003). From the Statistics of Data to the Statistics of Knowledge: Symbolic Data Analysis. Journal of the American of the Statistical Association, USA.

See Also

`read.sym.table`

Examples

```
data(example3)
display.sym.table(example3)
```

ex1_db2so *Data example to generate symbolic objets*

Description

This is a small data example to generate symbolic objets.

Usage

```
data(ex1_db2so)
```

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Examples

```
data(ex1_db2so)
ex1 <- ex1_db2so
result <- classic.to.sym(ex1, concept=c("state", "sex"),
                    variables=c("county", "group", "age","age"),
                    variables.types=c("$C", "$I", "$H", "$S"))
result
```

example1 *Data Example 1*

Description

This a symbolic data table with variables of continuos, interval, histogram and set types.

Usage

```
data(example1)
```

Format

The labels $C means that follows a continuous variable, $I means an interval variable, $H means a histogram variables and $S means set variable. In the first row each labels should be follow of a name to variable and to the case of histogram a set variables types the names of the modalities (categories). In data rows for continuous variables we have just one value, for interval variables we have the minimum and the maximum of the interval, for histogram variables we have the number of modalities and then the probability of each modality and for set variables we have the cardinality of the set and next the elements of the set.

The format is the *.csv file is:

$C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4
Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i
Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d
Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c
Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a
Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k
The internal format is:
$N
[1] 5
$M
[1] 4
$sym.obj.names
[1] "Case1" "Case2" "Case3" "Case4" "Case5"
$sym.var.names
[1] "F1" "F2" "F3" "F4"
$sym.var.types
[1] "$C" "$I" "$H" "$S"
$sym.var.length
[1] 1 2 3 4
$sym.var.starts
[1] 2 4 8 13
$meta
$C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4
Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i
Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d
Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c
Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a
Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k
$data
F1 F2 F2.1 M1 M2 M3 E1 E2 E3 E4
Case1 2.8 1 2 0.1 0.7 0.2 e g k i
Case2 1.4 3 9 0.6 0.3 0.1 a b c d
Case3 3.2 -1 4 0.2 0.2 0.6 2 1 b c
Case4 -2.1 0 2 0.9 0.0 0.1 3 4 c a
Case5 -3.0 -4 -2 0.6 0.0 0.4 e i g k

Examples

```
data(example1)
```

example2 *Data Example 2*

Description

This a symbolic data matrix wint continuos, interval, histograma a set data types.

Usage

```
data(example2)
```

Format

$C F1 $I F2 F2 $H F3 M1 M2 M3 $C F4 $S F5 E1 E2 E3 E4

Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $C 6.0 $S 4 e g k i

Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $C 8.0 $S 4 a b c d

Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $C -7.0 $S 4 2 1 b c

Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $C 0.0 $S 4 3 4 c a

Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $C -9.5 $S 4 e i g k

Examples

```
data(example2)
display.sym.table(example2)
```

example3 *Data Example 3*

Description

This a symbolic data matrix wint continuos, interval, histograma a set data types.

Usage

```
data(example3)
```

Format

$C F1 $I F2 F2 $H F3 M1 M2 M3 $C F4 $S F5 E1 E2 E3 E4 $I F6 F6 $I F7 F7

Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $C 6.0 $S 4 e g k i $I 0.00 90.00 $I 9 24

Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $C 8.0 $S 4 a b c d $I -90.00 98.00 $I -9 9

Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $C -7.0 $S 4 2 1 b c $I 65.00 90.00 $I 65 70

Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $C 0.0 $S 4 3 4 c a $I 45.00 89.00 $I 25 67

Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $C -9.5 $S 4 e i g k $I 20.00 40.00 $I 9 40

Case6 $C 0.1 $I 10 21 $H 3 0.0 0.7 0.3 $C -1.0 $S 4 f t c a $I 5.00 8.00 $I 5 8

Case7 $C 9.0 $I 4 21 $H 3 0.2 0.2 0.6 $C 0.5 $S 4 e q r z $I 3.14 6.76 $I 4 6

Examples

```
data(example3)
display.sym.table(example3)
```

---

example4 *Data Example 4*

Description

data(example4) display.sym.table(example4)

Usage

```
data(example4)
```

Format

$C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $C 6 $S 4 e g k i $I 0 90

Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $C 8.0 $S 4 a b c d $I -90.00 98.00

Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $C -7.0 $S 4 2 1 b c $I 65.00 90.00

Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $C 0.0 $S 4 3 4 c a $I 45.00 89.00

Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $C -9.5 $S 4 e i g k $I 90.00 990.00

Case6 $C 0.1 $I 10 21 $H 3 0.0 0.7 0.3 $C -1.0 $S 4 f t c a $I 5.00 8.00

Case7 $C 9.0 $I 4 21 $H 3 0.2 0.2 0.6 $C 0.5 $S 4 e q r z $I 3.14 6.76

Examples

```
data(example4)
display.sym.table(example4)
```

example5 *Data Example 5*

Description

This a symbolic data matrix wint continuos, interval, histograma a set data types.

Usage

```
data(example5)
```

Format

$H F0 M01 M02 $C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4

Case1 $H 2 0.1 0.9 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i

Case2 $H 2 0.7 0.3 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d

Case3 $H 2 0.0 1.0 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c

Case4 $H 2 0.2 0.8 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a

Case5 $H 2 0.6 0.4 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k

Examples

```
data(example5)
display.sym.table(example5)
```

example6 *Data Example 7*

Description

This a symbolic data matrix wint continuos, interval, histograma a set data types.

Usage

```
data(example6)
```

Format

$C F1 $H F2 M1 M2 M3 M4 M5 $I F3 F3 $H F4 M1 M2 M3 $C F5 $S F6 E1 E2 E3 E4

Case1 $C 2.8 $H 5 0.1 0.1 0.1 0.1 0.6 $I 1 2 $H 3 0.1 0.7 0.2 $C 6.0 $S 4 e g k i

Case2 $C 1.4 $H 5 0.1 0.1 0.1 0.1 0.6 $I 3 9 $H 3 0.6 0.3 0.1 $C 8.0 $S 4 a b c d

Case3 $C 3.2 $H 5 0.1 0.1 0.1 0.1 0.6 $I -1 4 $H 3 0.2 0.2 0.6 $C -7.0 $S 4 2 1 b c

Case4 $C -2.1 $H 5 0.1 0.1 0.1 0.1 0.6 $I 0 2 $H 3 0.9 0.0 0.1 $C 0.0 $S 4 3 4 c a

Case5 $C -3.0 $H 5 0.1 0.1 0.1 0.1 0.6 $I -4 -2 $H 3 0.6 0.0 0.4 $C -9.5 $S 4 e i g k

Examples

```
data(example6)
display.sym.table(example6)
```

---

example7 *Data Example 7*

Description

This a symbolic data matrix wint continuos, interval, histograma a set data types.

Usage

```
data(example7)
```

Format

$C F1 $H F2 M1 M2 M3 M4 M5 $I F3 F3 $H F4 M1 M2 M3 $C F5

Case1 $C 2.8 $H 5 0.1 0.2 0.3 0.4 0.0 $I 1 2 $H 3 0.1 0.7 0.2 $C 6.0

Case2 $C 1.4 $H 5 0.2 0.1 0.5 0.1 0.2 $I 3 9 $H 3 0.6 0.3 0.1 $C 8.0

Case3 $C 3.2 $H 5 0.1 0.1 0.2 0.1 0.5 $I -1 4 $H 3 0.2 0.2 0.6 $C -7.0

Case4 $C -2.1 $H 5 0.4 0.1 0.1 0.1 0.3 $I 0 2 $H 3 0.9 0.0 0.1 $C 0.0

Case5 $C -3.0 $H 5 0.6 0.1 0.1 0.1 0.1 $I -4 -2 $H 3 0.6 0.0 0.4 $C -9.5

Examples

```
data(example7)
display.sym.table(example7)
```

---

ex_cfa1 *Correspondence Analysis Example*

Description

Correspondence Analysis for Symbolic MultiValued Variables example.

Usage

```
data(ex_cfa1)
```

References

Rodriguez, O. (2011). Correspondence Analysis for Symbolic MultiValued Variables. Workshop in Symbolic Data Analysis Namur, Belgium

Examples

```
data(ex_cfa1)
res<-sym.cfa(ex_cfa1)
cfa.scatterplot(sym.var(res,1),sym.var(res,2),num.gr1=ex_cfa1$N,labels=TRUE,col='red',
                main='CFA')
```

---

ex_cfa2 *Correspondence Analysis Example*

Description

Correspondence Analysis for Symbolic MultiValued Variables example.

Usage

```
data(ex_cfa2)
```

References

Rodriguez, O. (2011). Correspondence Analysis for Symbolic MultiValued Variables. Workshop in Symbolic Data Analysis Namur, Belgium

Examples

```
data(ex_cfa2)
res<-sym.cfa(ex_cfa2)
cfa.scatterplot(sym.var(res,1),sym.var(res,2),num.gr1=ex_cfa2$N,labels=TRUE,col='red',
                main='CFA')
```

---

generate.sym.table *Generate a Symbolic Data Table*

Description

This function generates a symbolic data table from a CSV data file.

Usage

```
generate.sym.table(sym.data, file, sep, dec, row.names = NULL, col.names = NULL)
```

Arguments

sym.data	Symbolic data table
file	The name of the CSV file.
sep	As in R function read.table
dec	As in R function read.table
row.names	As in R function read.table
col.names	As in R function read.table

Value

Return a symbolic data table.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Examples

```
data(example1)
generate.sym.table(example1, file='temp4.csv', sep='|',dec='.', row.names=TRUE,
                   col.names=TRUE)
ex1<-read.sym.table('temp4.csv', header=TRUE, sep='|',dec='.', row.names=1)
```

---

interscal *Interscal Method*

Description

Execute interscal Method.

Usage

```
interscal(sym.data)
```

Arguments

sym.data	The symbolic data matrix.

Value

The symbolic interval components.

Author(s)

Oldemar Rodriguez Rojas

References

Groenen, P.J.F., Winsberg, S., Rodriguez, O., Diday, E. (2006). I-Scal: Multidimensional scaling of interval dissimilarities. Computational Statistics and Data Analysis, 51, 360-378.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University

See Also

sym.interval.pca

Examples

```
data(ex_cfa1)
res<-interscal(ex_cfa1)
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                labels=TRUE,col='red',main='Interscal CFA Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                  sym.var(res$Sym.Components,3),color='blue',
                  labels=TRUE,main='Interscal CFA Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                                    labels=TRUE)
```

---

interval.dist *Interval Distance Matrix*

Description

Compute a distance matrix from a symbolic interval data matrix.

Usage

```
interval.dist(sym.data, distance = c("hausdorff", "centers", "interscal"), p = 2)
```

Arguments

sym.data	Symbolic data matrix with the variables of interval type.
distance	The distance to be use.
p	The p in the Hausdorff distance

$$d(w_{u_1}, w_{u_2}) = \left(\sum_{j=1}^{m} \Phi_j(w_{u_1}, w_{u_2})^p \right)^{1/p}$$

Value

Return a R distance triangular matrix

Author(s)

Oldemar Rodriguez Rojas

References

Groenen, P.J.F., Winsberg, S., Rodriguez, O., Diday, E. (2006). I-Scal: Multidimensional scaling of interval dissimilarities. Computational Statistics and Data Analysis, 51, 360-378.

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Examples

```
data(VeterinaryData)
VD<-VeterinaryData
interval.dist(VD)
interval.dist(VD,distance='centers')
```

`interval.dist.tobj` *Symbolic Objects Distance*

Description

Compute a distance between two symbolic objects.

Usage

```
interval.dist.tobj(sym.obj.x, sym.obj.y, distance = c("hausdorff",
                           "centers", "interscal"), p = 2)
```

Arguments

`sym.obj.x`	First Symbolic Object
`sym.obj.y`	Second Symbolic Object
`distance`	Dsitance to be use
`p`	The p in the Hausdorff distance

$$d(w_{u_1}, w_{u_2}) = \left(\sum_{j=1}^{m} \Phi_j(w_{u_1}, w_{u_2})^p \right)^{1/p}$$

Value

Return a real number that is the distance between sym.obj.x and sym.obj.y

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

See Also

interval.dist

Examples

```
data(VeterinaryData)
VD<-VeterinaryData
interval.dist.tobj(sym.obj(VD,1),sym.obj(VD,2))
interval.dist.tobj(sym.obj(VD,1),sym.obj(VD,2),distance="centers")
```

---

int_prost_test *Linear regression model data example*

Description

Linear regression model interval-valued data example.

Usage

```
data(int_prost_test)
```

References

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_test,method='cm')
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm$Fitted)
RMSE.U(sym.var(int_prost_test,9),pred.cm$Fitted)
R2.L(sym.var(int_prost_test,9),pred.cm$Fitted)
R2.U(sym.var(int_prost_test,9),pred.cm$Fitted)
deter.coefficient(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
plot(res.cm.lasso)
plot(res.cm.lasso$glmnet.fit, "norm", label=TRUE)
plot(res.cm.lasso$glmnet.fit, "lambda", label=TRUE)
```

```
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm.lasso)
RMSE.U(sym.var(int_prost_test,9),pred.cm.lasso)
R2.L(sym.var(int_prost_test,9),pred.cm.lasso)
R2.U(sym.var(int_prost_test,9),pred.cm.lasso)
deter.coefficient(sym.var(int_prost_test,9),pred.cm.lasso)
```

`int_prost_train` *Linear regression model data example*

Description

Linear regression model interval-valued data example.

Usage

```
data(int_prost_train)
```

References

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_test,method='cm')
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm$Fitted)
RMSE.U(sym.var(int_prost_test,9),pred.cm$Fitted)
R2.L(sym.var(int_prost_test,9),pred.cm$Fitted)
R2.U(sym.var(int_prost_test,9),pred.cm$Fitted)
deter.coefficient(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
plot(res.cm.lasso)
plot(res.cm.lasso$glmnet.fit, "norm", label=TRUE)
plot(res.cm.lasso$glmnet.fit, "lambda", label=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm.lasso)
RMSE.U(sym.var(int_prost_test,9),pred.cm.lasso)
R2.L(sym.var(int_prost_test,9),pred.cm.lasso)
R2.U(sym.var(int_prost_test,9),pred.cm.lasso)
deter.coefficient(sym.var(int_prost_test,9),pred.cm.lasso)
```

lynne1 *Symbolic interval data example*

Description

Symbolic data matrix with all the variables of interval type

Usage

```
data(lynne1)
```

Format

$I Pulse Rate Pulse Rate $I Systolic Pressure Systolic Pressure $I Diastolic Pressure

1 $I 44 68 $I 90 110 $I 50

2 $I 60 72 $I 90 130 $I 70

3 $I 56 90 $I 140 180 $I 90

4 $I 70 112 $I 110 142 $I 80

5 $I 54 72 $I 90 100 $I 50

6 $I 70 100 $I 134 142 $I 80

7 $I 72 100 $I 130 160 $I 76

8 $I 76 98 $I 110 190 $I 70

9 $I 86 96 $I 138 180 $I 90

10 $I 86 100 $I 110 150 $I 28

Diastolic Pressure

1 70

2 90

3 100

4 108

5 70

6 110

7 90

8 110

9 110

10 100

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Examples

```
data(lynne1)
display.sym.table(lynne1)
```

---

lynne2 *Symbolic interval data example*

Description

Symbolic data matrix with all the variables of interval type

Usage

```
data(lynne2)
```

Format

$I Pulse Rate Pulse Rate $I Systolic Pressure Systolic Pressure $I Diastolic Pressure

1 $I 44 68 $I 90 110 $I 50

2 $I 60 72 $I 90 130 $I 70

3 $I 56 90 $I 140 180 $I 90

4 $I 70 112 $I 110 142 $I 80

5 $I 54 72 $I 90 100 $I 50

6 $I 70 100 $I 134 142 $I 80

7 $I 72 100 $I 130 160 $I 76

8 $I 76 98 $I 110 190 $I 70

9 $I 86 96 $I 138 180 $I 90

10 $I 86 100 $I 110 150 $I 78

11 $I 53 55 $I 160 190 $I 205

12 $I 50 55 $I 180 200 $I 110

13 $I 73 81 $I 125 138 $I 78

14 $I 60 75 $I 175 194 $I 90

15 $I 42 52 $I 105 115 $I 70

Diastolic Pressure

1 70

2 90

3 100

4 108

5 70

6 110

7 90

8 110

9 110

10 100

11 219

12 125

13 99

14 100

15 82

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Examples

```
data(lynne2)
display.sym.table(lynne2)
```

`oils` *Ichino Oils example data*

Description

Symbolic data matrix with all the variables of interval type

Usage

```
data(oils)
```

Format

$I GRA GRA $I FRE FRE $I IOD IOD $I SAP SAP

L $I 0.930 0.935 $I -27 -18 $I 170 204 $I 118 196

P $I 0.930 0.937 $I -5 -4 $I 192 208 $I 188 197

Co $I 0.916 0.918 $I -6 -1 $I 99 113 $I 189 198

S $I 0.920 0.926 $I -6 -4 $I 104 116 $I 187 193

Ca $I 0.916 0.917 $I -25 -15 $I 80 82 $I 189 193

O $I 0.914 0.919 $I 0 6 $I 79 90 $I 187 196

B $I 0.860 0.870 $I 30 38 $I 40 48 $I 190 199

H $I 0.858 0.864 $I 22 32 $I 53 77 $I 190 202

References

Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales a des donnees de type intervalle, Rev. Statistique Appliquee, Vol. XLV Num. 3 pag. 5-24, France.

Examples

```
data(oils)
display.sym.table(oils)
```

`predictsym.glm` *Predict method to Lasso, Ridge and and Elastic Net Linear regression model to interval variables*

Description

To execute Predict method to Lasso, Ridge and and Elastic Net Linear regression model to interval variables.

Usage

```
predictsym.glm(model, new.sym.data, response = 1, method = c("cm", "crm"))
```

Arguments

`model`	The output of glm method.
`new.sym.data`	Should be a symbolic data table read with the function read.sym.table(...).
`response`	The number of the column where is the response variable in the interval data table.
`method`	"cm" to generalized Center Method and "crm" to generalized Center and Range Method.

Value

The object returned depends the ... argument which is passed on to the predict method for glmnet objects.

Author(s)

Oldemar Rodriguez Rojas

References

Rodriguez O. (2013). A generalization of Centre and Range method for fitting a linear regression model to symbolic interval data using Ridge Regression, Lasso and Elastic Net methods. The IFCS2013 conference of the International Federation of Classification Societies, Tilburg University Holland.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
plot(res.cm.lasso)
plot(res.cm.lasso$glmnet.fit, "norm", label=TRUE)
plot(res.cm.lasso$glmnet.fit, "lambda", label=TRUE)
RMSE.L(sym.var(int_prost_test,9),pred.cm.lasso)
RMSE.U(sym.var(int_prost_test,9),pred.cm.lasso)
R2.L(sym.var(int_prost_test,9),pred.cm.lasso)
R2.U(sym.var(int_prost_test,9),pred.cm.lasso)
deter.coefficient(sym.var(int_prost_test,9),pred.cm.lasso)
```

---

predictsym.lm *Predict method to CM and CRM Linear regression model*

Description

To execute predict method the Center Method (CR) and Center and Range Method (CRM) to Linear regression

Usage

```
predictsym.lm(model, new.sym.data, method = c("cm", "crm"))
```

Arguments

model	The output of lm method.
new.sym.data	Should be a symbolic data table read with the function read.sym.table(...).
method	"cm" to Center Method and "crm" to Center and Range Method.

Value

predictsym.lm produces a vector of predictions or a matrix of predictions and bounds with column names fit, lwr, and upr if interval is set. For type = "terms" this is a matrix with a column per term and may have an attribute "constant"

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
```

R2.L *Lower boundary correlation coefficient.*

Description

Compute the lower boundary correlation coefficient for two interval variables.

Usage

```
R2.L(sym.var, prediction)
```

Arguments

sym.var	Variable that was predicted.
prediction	The prediction given by the model.

Value

The lower boundary correlation coefficient.

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
R2.L(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
R2.L(sym.var(int_prost_test,9),pred.cm.lasso)
```

R2.U *Upper boundary correlation coefficient.*

Description

Compute the upper boundary correlation coefficient for two interval variables.

Usage

```
R2.U(sym.var, prediction)
```

Arguments

sym.var	Variable that was predicted.
prediction	The prediction given by the model.

Value

The upper boundary correlation coefficient.

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
R2.U(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
R2.U(sym.var(int_prost_test,9),pred.cm.lasso)
```

---

read.sym.table *Read a Symbolic Table*

Description

It reads a symbolic data table from a CSV file.

Usage

```
read.sym.table(file, header = TRUE, sep, dec, row.names = NULL)
```

Arguments

file	The name of the CSV file.
header	As in R function read.table
sep	As in R function read.table
dec	As in R function read.table
row.names	As in R function read.table

Details

The labels $C means that follows a continuous variable, $I means an interval variable, $H means a histogram variables and $S means set variable. In the first row each labels should be follow of a name to variable and to the case of histogram a set variables types the names of the modalities (categories) . In data rows for continuous variables we have just one value, for interval variables we have the minimum and the maximum of the interval, for histogram variables we have the number of modalities and then the probability of each modality and for set variables we have the cardinality of the set and next the elements of the set.

The format is the CSV file should be like:

$C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4

Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i
Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d
Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c
Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a
Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k
The internal format is:
$N
[1] 5
$M
[1] 4
$sym.obj.names
[1] "Case1" "Case2" "Case3" "Case4" "Case5"
$sym.var.names
[1] "F1" "F2" "F3" "F4"
$sym.var.types
[1] "$C" "$I" "$H" "$S"
$sym.var.length
[1] 1 2 3 4
$sym.var.starts
[1] 2 4 8 13
$meta
$C F1 $I F2 F2 $H F3 M1 M2 M3 $S F4 E1 E2 E3 E4
Case1 $C 2.8 $I 1 2 $H 3 0.1 0.7 0.2 $S 4 e g k i
Case2 $C 1.4 $I 3 9 $H 3 0.6 0.3 0.1 $S 4 a b c d
Case3 $C 3.2 $I -1 4 $H 3 0.2 0.2 0.6 $S 4 2 1 b c
Case4 $C -2.1 $I 0 2 $H 3 0.9 0.0 0.1 $S 4 3 4 c a
Case5 $C -3.0 $I -4 -2 $H 3 0.6 0.0 0.4 $S 4 e i g k
$data
F1 F2 F2.1 M1 M2 M3 E1 E2 E3 E4
Case1 2.8 1 2 0.1 0.7 0.2 e g k i
Case2 1.4 3 9 0.6 0.3 0.1 a b c d
Case3 3.2 -1 4 0.2 0.2 0.6 2 1 b c
Case4 -2.1 0 2 0.9 0.0 0.1 3 4 c a
Case5 -3.0 -4 -2 0.6 0.0 0.4 e i g k

Value

Return a symbolic data table structure.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

display.sym.table

Examples

```
data(example1)
write.sym.table(example1, file='temp4.csv', sep='|',dec='.', row.names=TRUE,
                col.names=TRUE)
ex1<-read.sym.table('temp4.csv', header=TRUE, sep='|',dec='.', row.names=1)
```

RMSE.L *Lower boundary root-mean-square error*

Description

Compute the lower boundary root-mean-square error

Usage

```
RMSE.L(sym.var, prediction)
```

Arguments

sym.var	Variable that was predicted.
prediction	The prediction given by the model.

Value

The lower boundary root-mean-square error.

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm.lasso)
```

RMSE.U *Upper boundary root-mean-square error*

Description

Compute the upper boundary root-mean-square error

Usage

```
RMSE.U(sym.var, prediction)
```

Arguments

sym.var	Variable that was predicted.
prediction	The prediction given by the model.

Value

The upper boundary root-mean-square error.

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
RMSE.U(sym.var(int_prost_test,9),pred.cm$Fitted)

res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
RMSE.U(sym.var(int_prost_test,9),pred.cm.lasso)
```

SODAS.to.RSDA *XML SODAS files to RSDA files.*

Description

To convert XML SODAS files to RSDA files.

Usage

```
SODAS.to.RSDA(XMLPath, labels = T)
```

Arguments

XMLPath	Disk path where the SODAS file is.
labels	If we want to include SODAS files lebels in RSDA file.

Value

A RSDA symbolic data file.

Author(s)

Olger Calderon and Roberto Zuniga.

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

abalone

Examples

```
# We can read the file directly from the SODAS XML file as follows:
# abalone<-SODAS.to.RSDA("C:/Program Files (x86)/DECISIA/SODAS version 2.0/bases/abalone.xml)
# We can save the file in CSV to RSDA format as follows:
# write.sym.table(sodas.ex1, file='abalone.csv', sep=';',dec='.', row.names=TRUE,
#               col.names=TRUE)
# We read the file from the CSV file,
# this is not necessary if the file is read directly from
# XML using SODAS.to.RSDA as in the first statement in this example.
data(abalone)
res<-sym.interval.pca(abalone,'centers')
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                labels=TRUE,col='red',main='PCA Oils Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                  sym.var(res$Sym.Components,3),color='blue',main='PCA Oils Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                       labels=TRUE)
sym.circle.plot(res$Sym.Prin.Correlations)
```

StudentsGrades *Data Example*

Description

Symbolic data matrix with all the variables continuos type.

Usage

```
data(StudentsGrades)
```

Format

$C Math $C Science $C Spanish $C History $C Sport

Lucia $C 7.0 $C 6.5 $C 9.2 $C 8.6 $C 8.0

Pedro $C 7.5 $C 9.4 $C 7.3 $C 7.0 $C 7.0

Ines $C 7.6 $C 9.2 $C 8.0 $C 8.0 $C 7.5

Luis $C 5.0 $C 6.5 $C 6.5 $C 7.0 $C 9.0

Andres $C 6.0 $C 6.0 $C 7.8 $C 8.9 $C 7.3

Ana $C 7.8 $C 9.6 $C 7.7 $C 8.0 $C 6.5

Carlos $C 6.3 $C 6.4 $C 8.2 $C 9.0 $C 7.2

Jose $C 7.9 $C 9.7 $C 7.5 $C 8.0 $C 6.0

Sonia $C 6.0 $C 6.0 $C 6.5 $C 5.5 $C 8.7

Maria $C 6.8 $C 7.2 $C 8.7 $C 9.0 $C 7.0

Examples

```
data(StudentsGrades)
StudentsGrades
```

`sym.cfa` *Symbolic correspondence analysis*

Description

Correspondence Analysis for Symbolic MultiValued Variables.

Usage

```
sym.cfa(sym.data)
```

Arguments

`sym.data`	Should be a symbolic data table read with the function read.sym.table(...).

Value

Return the interval principal components.

Author(s)

Oldemar Rodriguez Rojas

References

Rodriguez, O. (2011). Correspondence Analysis for Symbolic MultiValued Variables. Workshop in Symbolic Data Analysis Namur, Belgium.

Examples

```
data(ex_cfa1)
res<-sym.cfa(ex_cfa1)
cfa.scatterplot(sym.var(res,1),sym.var(res,2),num.gr1=ex_cfa1$N,
                        labels=TRUE,col='red',main='CFA')
```

sym.circle.plot *Symbolic Circle of Correlations*

Description

Plot the symbolic circle of correlations

Usage

```
sym.circle.plot(prin.corre)
```

Arguments

prin.corre	A symbolic interval data matrix with correlations between the variables and the principals componets, both of interval type.

Value

Plot the symbolic circle

Author(s)

Oldemar Rodriguez Rojas

References

Rodriguez O. (2012). The Duality Problem in Interval Principal Components Analysis. The 3rd Workshop in Symbolic Data Analysis, Madrid.

Examples

```
data(oils)
res<-sym.interval.pca(oils,'centers')
sym.circle.plot(res$Sym.Prin.Correlations)
```

sym.cor *Symbolic Correlation*

Description

This function compute the symbolic correlation

Usage

```
sym.cor(sym.var.x, sym.var.y, method =
          c("centers", "interval", "billard", "histogram"), na.rm = FALSE, ...)
```

Arguments

sym.var.x	First symbolic variables.
sym.var.y	Second symbolic variables.
method	The method to be use.
na.rm	As in R cor function.
...	As in R cor function.

Value

Return a real number in [-1,1].

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
sym.cor(sym.var(sym.data,1),sym.var(sym.data,4),method='centers')
sym.cor(sym.var(sym.data,2),sym.var(sym.data,6),method='centers')
sym.cor(sym.var(sym.data,2),sym.var(sym.data,6),method='billard')
```

---

sym.cov *Symbolic Covariance*

Description

This function compute the symbolic covariance

Usage

```
sym.cov(sym.var.x, sym.var.y, method = c("centers",
              "interval", "billard", "histogram"), na.rm = FALSE, ...)
```

Arguments

sym.var.x	First symbolic variables.
sym.var.y	Second symbolic variables.
method	The method to be use.
na.rm	As in R cov function.
...	As in R cov function.

Value

Return a real number.

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
sym.cov(sym.var(sym.data,1),sym.var(sym.data,4),method='centers')
sym.cov(sym.var(sym.data,2),sym.var(sym.data,6),method='centers')
sym.cov(sym.var(sym.data,2),sym.var(sym.data,6),method='billard')
```

---

sym.glm — *Lasso, Ridge and and Elastic Net Linear regression model to interval variables*

Description

Execute Lasso, Ridge and and Elastic Net Linear regression model to interval variables.

Usage

```
sym.glm(sym.data, response = 1, method = c("cm", "crm"),
               alpha = 1, nfolds = 10, grouped = TRUE)
```

Arguments

sym.data	Should be a symbolic data table read with the function read.sym.table(...).
response	The number of the column where is the response variable in the interval data table.
method	"cm" to generalized Center Method and "crm" to generalized Center and Range Method.
alpha	alpha=1 is the lasso penalty, and alpha=0 the ridge penalty. 0<alpha<1 is the elastic net method.
nfolds	Number of folds - default is 10. Although nfolds can be as large as the sample size (leave-one-out CV), it is not recommended for large datasets. Smallest value allowable is nfolds=3
grouped	This is an experimental argument, with default TRUE, and can be ignored by most users.

Value

An object of class "cv.glmnet" is returned, which is a list with the ingredients of the cross-validation fit.

Author(s)

Oldemar Rodriguez Rojas

References

Rodriguez O. (2013). A generalization of Centre and Range method for fitting a linear regression model to symbolic interval data using Ridge Regression, Lasso and Elastic Net methods. The IFCS2013 conference of the International Federation of Classification Societies, Tilburg University Holland.

See Also

sym.lm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm.lasso<-sym.glm(sym.data=int_prost_train,response=9,method='cm',
                      alpha=1,nfolds=10,grouped=TRUE)
pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=9,int_prost_test,method='cm')
plot(res.cm.lasso)
plot(res.cm.lasso$glmnet.fit, "norm", label=TRUE)
plot(res.cm.lasso$glmnet.fit, "lambda", label=TRUE)
RMSE.L(sym.var(int_prost_test,9),pred.cm.lasso)
RMSE.U(sym.var(int_prost_test,9),pred.cm.lasso)
R2.L(sym.var(int_prost_test,9),pred.cm.lasso)
R2.U(sym.var(int_prost_test,9),pred.cm.lasso)
deter.coefficient(sym.var(int_prost_test,9),pred.cm.lasso)
```

sym.hclust *Symbolic Hierarchical Clustering*

Description

This function allows us to execute a symbolic hierarchical clustering to interval variables.

Usage

```
sym.hclust(sym.data, distance = c("hausdorff", "centers"), p = 2,
                 method = c("ward", "single", "complete", "average", "mcquitty",
                   "median", "centroid"), members = NULL)
```

Arguments

sym.data	The symbolic data table.
distance	The distance to be use.
p	The p in the Hausdorff distance

$$d(w_{u_1}, w_{u_2}) = \left(\sum_{j=1}^{m} \Phi_j(w_{u_1}, w_{u_2})^p \right)^{1/p}$$

method	The method to be use, like in hclust R function.
members	Like in hclust R function.

Value

Return a dendogram plot structure.

Author(s)

Oldemar Rodriguez Rojas

References

Carvalho F., Souza R.,Chavent M., and Lechevallier Y. (2006) Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. Pattern Recognition Letters Volume 27, Issue 3, February 2006, Pages 167-179

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(oils)
sh<-sym.hclust(oils)
plot(sh)
sh<-sym.hclust(oils,'centers')
plot(sh)
```

`sym.histogram.pca` *Histogram Principal Components Analysis*

Description

This functions allows us to execute a histogram principal components analysis from a symbolic data table with continuos, interval or histogram variables that can be mixed.

Usage

```
sym.histogram.pca(sym.data, method = c("histogram", "classic"))
```

Arguments

`sym.data`	Symbolic data table.
`method`	The method to be used.

Value

Return a symbolic data table.

Author(s)

Oldemar Rodriguez Rojas

References

Diday, E., Rodriguez O. and Winberg S. (2000). Generalization of the Principal Components Analysis to Histogram Data, 4th European Conference on Principles and Practice of Knowledge Discovery in Data Bases, September 12-16, 2000, Lyon, France.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

See Also

`sym.interval.pca`

Examples

```
data(example7)
res<-sym.histogram.pca(example7)
sym.scatterplot(sym.var(res,1),sym.var(res,2),labels=TRUE,col='red',main='Histogram PCA')
sym.scatterplot3d(sym.var(res,1),sym.var(res,2),sym.var(res,3),color='blue',
                  main='Histogram PCA')
```

`sym.interval.pca` *Interval Principal Components Analysis*

Description

Cazes, Chouakria, Diday and Schektman (1997) proposed the Centers and the Tops Methods to extend the well known principal components analysis method to a particular kind of symbolic objects characterized by multi–values variables of interval type.

Usage

```
sym.interval.pca(sym.data, method = c("classic", "tops", "centers"))
```

Arguments

`sym.data`	Shoud be a symbolic data table read with the function read.sym.table(...)
`method`	It is use so select the method, "classic" execute a classical principal component analysis over the centers of the intervals, "tops" to use the vertices algorithm and "centers" to use the centers algorithm.

Value

Sym.Components: This a symbolic data table with the interval principal components. As this is a symbolic data table we can apply over this table any other symbolic data analysis method (symbolic propagation).

Sym.Prin.Correlations: This is the interval correlations between the original interval variables and the interval principal components, it can be use to plot the symbolic circle of correlations.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Cazes P., Chouakria A., Diday E. et Schektman Y. (1997). Extension de l'analyse en composantes principales a des donnees de type intervalle, Rev. Statistique Appliquee, Vol. XLV Num. 3 pag. 5-24, France.

Chouakria A. (1998) Extension des methodes d'analysis factorialle a des donnees de type intervalle, Ph.D. Thesis, Paris IX Dauphine University.

Makosso-Kallyth S. and Diday E. (2012). Adaptation of interval PCA to symbolic histogram variables, Advances in Data Analysis and Classification July, Volume 6, Issue 2, pp 147-159.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

See Also

sym.histogram.pca

Examples

```
data(oils)
res<-sym.interval.pca(oils,'centers')
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                labels=TRUE,col='red',main='PCA Oils Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                  sym.var(res$Sym.Components,3),color='blue',main='PCA Oils Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                       labels=TRUE)
sym.circle.plot(res$Sym.Prin.Correlations)

res<-sym.interval.pca(oils,'classic')
plot.PCA(res,choix="ind")
plot.PCA(res,choix="var")

data(lynne2)
res<-sym.interval.pca(lynne2,'centers')
sym.scatterplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                labels=TRUE,col='red',main='PCA Lynne Data')
sym.scatterplot3d(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                  sym.var(res$Sym.Components,3),color='blue', main='PCA Lynne Data')
sym.scatterplot.ggplot(sym.var(res$Sym.Components,1),sym.var(res$Sym.Components,2),
                       labels=TRUE)
sym.circle.plot(res$Sym.Prin.Correlations)

data(StudentsGrades)
st<-StudentsGrades
s.pca<-sym.interval.pca(st)
plot.PCA(s.pca,choix="ind")
plot.PCA(s.pca,choix="var")
```

---

sym.kmeans *Symbolic k-Means*

Description

This is a function is to carry out a k-means overs a interval symbolic data matrix.

Usage

```
sym.kmeans(sym.data, k = 3, iter.max = 10, nstart = 1,
           algorithm = c("Hartigan-Wong", "Lloyd", "Forgy", "MacQueen"))
```

Arguments

sym.data	Symbolic data table.
k	The number of clusters.
iter.max	Maximun number of iterations.
nstart	As in R kmeans function.
algorithm	The method to be use, as in kmeans R function.

Value

This function return the following information:

K-means clustering with 3 clusters of sizes 2, 2, 4

Cluster means:

GRA FRE IOD SAP

1 0.93300 -13.500 193.500 174.75

2 0.86300 30.500 54.500 195.25

3 0.91825 -6.375 95.375 191.50

Clustering vector:

L P Co S Ca O B H

1 1 3 3 3 3 2 2

Within cluster sum of squares by cluster:

[1] 876.625 246.125 941.875

(between_SS / total_SS = 92.0

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss"

[7] "size"

Author(s)

Oldemar Rodriguez Rojas

References

Carvalho F., Souza R.,Chavent M., and Lechevallier Y. (2006) Adaptive Hausdorff distances and dynamic clustering of symbolic interval data. Pattern Recognition Letters Volume 27, Issue 3, February 2006, Pages 167-179

See Also

sym.hclust

Examples

```
data(oils)
sk<-sym.kmeans(oils,k=3)
sk$cluster
```

sym.lm *CM and CRM Linear regression model*

Description

To execute the Center Method (CR) and Center and Range Method (CRM) to Linear regression

Usage

```
sym.lm(formula, sym.data, method = c("cm", "crm"))
```

Arguments

formula	An object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
sym.data	Should be a symbolic data table read with the function read.sym.table(...).
method	"cm" to Center Method and "crm" to Center and Range Method.

Details

Models for lm are specified symbolically. A typical model has the form response ~ terms where response is the (numeric) response vector and terms is a series of terms which specifies a linear predictor for response. A terms specification of the form first + second indicates all the terms in first together with all the terms in second with duplicates removed. A specification of the form first:second indicates the set of terms obtained by taking the interactions of all terms in first with all terms in second. The specification first*second indicates the cross of first and second. This is the same as first + second + first:second.

Value

sym.lm returns an object of class "lm" or for multiple responses of class c("mlm", "lm")

Author(s)

Oldemar Rodriguez Rojas

References

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2008). Centre and range method to fitting a linear regression model on symbolic interval data. Computational Statistics and Data Analysis 52, 1500-1515.

LIMA-NETO, E.A., DE CARVALHO, F.A.T., (2010). Constrained linear regression models for symbolic interval-valued variables. Computational Statistics and Data Analysis 54, 333-347.

See Also

sym.glm

Examples

```
data(int_prost_train)
data(int_prost_test)
res.cm<-sym.lm(lpsa~.,sym.data=int_prost_train,method='cm')
pred.cm<-predictsym.lm(res.cm,int_prost_test,method='cm')
RMSE.L(sym.var(int_prost_test,9),pred.cm$Fitted)
RMSE.U(sym.var(int_prost_test,9),pred.cm$Fitted)
R2.L(sym.var(int_prost_test,9),pred.cm$Fitted)
R2.U(sym.var(int_prost_test,9),pred.cm$Fitted)
deter.coefficient(sym.var(int_prost_test,9),pred.cm$Fitted)
```

---

sym.lm.bi *Symbolic Linear Regression for two variables*

Description

The function build a symbolic regression for two interval or continuos variables.

Usage

```
sym.lm.bi(sym.var.x, sym.var.y, method = c("mid-points", "tops", "inf-sup",
                "billard"))
```

Arguments

sym.var.x	The firth symbolic variable.
sym.var.y	The second symbolic variable.
method	The thirth symbolic variable.

Value

This function return a regression structure as folllows:

$Intercept [1] 38.64236

$Beta1 [1] 0.3081313

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
lm.mod<-sym.lm.bi(sym.var(sym.data,1),sym.var(sym.data,4))
sym.scatterplot(sym.var(sym.data,1),sym.var(sym.data,4),col='blue',
                main='Linear Regression')
abline(lm.mod,lwd=3)

lm.mod<-sym.lm.bi(sym.var(sym.data,2),sym.var(sym.data,6))
sym.scatterplot(sym.var(sym.data,2),sym.var(sym.data,6),
col='blue',main='Linear Regression')
abline(lm.mod,lwd=3)

data(lynne1)
lm.mod<-sym.lm.bi(sym.var(lynne1,2),sym.var(lynne1,1))
sym.scatterplot(sym.var(lynne1,2),sym.var(lynne1,1),labels=TRUE,
col='red',main='Linear Regression')
abline(lm.mod,lwd=3,col='blue')

lm.mod<-sym.lm.bi(sym.var(lynne1,2),sym.var(lynne1,1),method='inf-sup')
sym.scatterplot(sym.var(lynne1,2),sym.var(lynne1,1),labels=TRUE,
col='red',main='Linear Regression')
abline(lm.mod$inf,lwd=3,col='blue')
abline(lm.mod$sup,lwd=3,col='blue')

lm.mod<-sym.lm.bi(sym.var(lynne1,2),sym.var(lynne1,1),method='tops')
sym.scatterplot(sym.var(lynne1,2),sym.var(lynne1,1),labels=TRUE,
col='red',main='Linear Regression')
abline(lm.mod,lwd=3,col='blue')

lm.mod<-sym.lm.bi(sym.var(lynne1,2),sym.var(lynne1,1),method='billard')
sym.scatterplot(sym.var(lynne1,2),sym.var(lynne1,1),labels=TRUE,
col='red',main='Linear Regression')
abline(lm.mod$Intercept,lm.mod$Beta1,lwd=3,col='blue')
```

---

sym.mds *Symbolic Multidemensional Scaling*

Description

This function execute a multidimensional scaling from a interval symbolic data matrix.

Usage

```
sym.mds(sym.data, distance = c("hausdorff", "centers"), p = 2,
        method = c("classic", "INTERSCAL"))
```

Arguments

sym.data	The symbolic data matrix.
distance	The distance to be use.
p	The p in the Hausdorff distance

$$d(w_{u_1}, w_{u_2}) = \left(\sum_{j=1}^{m} \Phi_j(w_{u_1}, w_{u_2})^p \right)^{1/p}$$

method	The method to be used.

Value

Return the coordanates to plot the graphic.

Author(s)

Oldemar Rodriguez Rojas

References

Groenen, P.J.F., Winsberg, S., Rodriguez, O., Diday, E. (2006). I-Scal: Multidimensional scaling of interval dissimilarities. Computational Statistics and Data Analysis, 51, 360-378.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

See Also

sym.interval.pca

Examples

```
data(oils)
res<-sym.mds(oils)
plot(res,pch = 23, bg = "red", xlab = "Score 1", ylab = "Score 2")
res<-sym.mds(oils,distance="centers")
plot(res,pch = 23, bg = "red", xlab = "Score 1", ylab = "Score 2")
```

---

sym.mean *Symbolic Mean*

Description

This function compute the symbolic mean

Usage

```
sym.mean(sym.var, method = c("centers", "interval", "histogram"),
         trim = 0, na.rm = FALSE, ...)
```

Arguments

sym.var	The symbolic variable.
method	The method to be use.
trim	As in R mean function.
na.rm	As in R mean function.
...	As in R mean function.

Value

Return a real number.

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
sym.mean(sym.var(sym.data,1))
sym.mean(sym.var(sym.data,2))
sym.mean(sym.var(sym.data,2),method='interval')
sym.mean(sym.var(sym.data,3),method='histogram')
```

---

sym.median *Symbolic Median*

Description

This function compute the symbolic median

Usage

```
sym.median(sym.var, method =
                   c("centers", "interval", "histogram"), na.rm = FALSE, ...)
```

Arguments

sym.var	The symbolic variable.
method	The method to be use.
na.rm	As in R median function.
...	As in R median function.

Value

Return a real number.

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
sym.median(sym.var(sym.data,1))
sym.median(sym.var(sym.data,2))
sym.median(sym.var(sym.data,6),method='interval')
sym.median(sym.var(sym.data,3),method='histogram')
```

---

sym.obj *Symbolic Object*

Description

This function get a symbolic object (row or a case) from a symbolic data table.

Usage

```
sym.obj(sym.data, number.sym.obj)
```

Arguments

sym.data	Symboli data matrix.
number.sym.obj	The number of the row for the symbolic object (case) that we want to get.

Value

Return a symbolic object with the following internal format:

$M

[1] 5

$var.types

[1] "$C" "$H" "$I" "$H" "$C"

$var.length

[1] 1 5 2 3 1

$var.names

[1] "F1" "F2" "F3" "F4" "F5"

$obj.data.vector

F1 M1 M2 M3 M4 M5 F3 F3.1 M1.1 M2.1 M3.1 F5

Case4 -2.1 0.4 0.1 0.1 0.1 0.3 0 2 0.9 0 0.1 0

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

sym.var

Examples

```
data(example7)
sym.obj(example7,4)
```

---

sym.scatterplot *Symbolic Scatter Plot*

Description

This function could be use to plot two symbolic variables in a X-Y plane.

Usage

```
sym.scatterplot(sym.var.x, sym.var.y, labels = FALSE, ...)
```

Arguments

sym.var.x	First symbolic variable
sym.var.y	Second symbolic variable.
labels	As in R plot function.
...	As in R plot function.

Value

Return a graphics.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

See Also

sym.scatterplot3d

Examples

```
data(example3)
sym.data<-example3
sym.scatterplot(sym.var(sym.data,1),sym.var(sym.data,4),col='blue',
                main='Main Title')
sym.scatterplot(sym.var(sym.data,1),sym.var(sym.data,4),labels=TRUE,col='blue',
                main='Main Title')
sym.scatterplot(sym.var(sym.data,2),sym.var(sym.data,6),labels=TRUE,
                col='red',main='Main Title',lwd=3)

data(oils)
sym.scatterplot(sym.var(oils,2),sym.var(oils,3),labels=TRUE,
                col='red',main='Oils Data')
data(lynne1)

sym.scatterplot(sym.var(lynne1,2),sym.var(lynne1,1),labels=TRUE,
                col='red',main='Lynne Data')
```

`sym.scatterplot.ggplot` *Symbolic Scatter GGPlot*

Description

This function could be use to plot two symbolic variables in a X-Y plane using ggplot R package.

Usage

```
sym.scatterplot.ggplot(sym.var.x, sym.var.y, labels = FALSE, ...)
```

Arguments

`sym.var.x`	First symbolic variable.
`sym.var.y`	Second symbolic variable.
`labels`	As in ggplot.
`...`	As in ggplot.

Value

return a ggplot graphic.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

See Also

`sym.scatterplot`

Examples

```
data(lynne1)
sym.scatterplot.ggplot(sym.var(lynne1,1),sym.var(lynne1,3),labels=TRUE)
data(oils)
sym.scatterplot.ggplot(sym.var(oils,2),sym.var(oils,3),labels=TRUE)
```

`sym.scatterplot3d` *Symbolic Scatter Plot 3D*

Description

This function could be use to plot two symbolic variables in 3D i.e. in a X-Y-Z plane.

Usage

```
sym.scatterplot3d(sym.var.x, sym.var.y, sym.var.z, labels = FALSE, ...)
```

Arguments

`sym.var.x`	First symbolic variable.
`sym.var.y`	Second symbolic variable.
`sym.var.z`	Third symbolic variable.
`labels`	As in R plot function.
`...`	As in R plot function.

Value

3D Plot graphic.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

`sym.scatterplot`

Examples

```
data(lynne1)
sym.scatterplot3d(sym.var(lynne1,1),sym.var(lynne1,2),sym.var(lynne1,3),
                            color='blue', main='Lynne Data')
```

sym.sd *Symbolic Standard Desviation*

Description

Compute the symbolic standard desviation.

Usage

```
sym.sd(sym.var, method =
          c("centers", "interval", "billard", "histogram"), na.rm = FALSE, ...)
```

Arguments

sym.var	The symbolic variable.
method	The method to be use.
na.rm	As in R sd function.
...	As in R sd function.

Value

return a real number.

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
sym.sd(sym.var(sym.data,1))
sym.sd(sym.var(sym.data,2))
sym.sd(sym.var(sym.data,6))
sym.sd(sym.var(sym.data,6),method='interval')
sym.sd(sym.var(sym.data,6),method='billard')
sym.sd(sym.var(sym.data,3),method='histogram')
```

sym.var *Symbolic Variable (Feature)*

Description

This function get a symbolic variable from a symbolic data table.

Usage

```
sym.var(sym.data, number.sym.var)
```

Arguments

sym.data	The symbolic data table
number.sym.var	The number of the column for the variable (feature) that we want to get.

Value

Return a symbolic data variable with the following structure:

$N

[1] 7

$var.name

[1] "F6"

$var.type

[1] "$I"

$obj.names

[1] "Case1" "Case2" "Case3" "Case4" "Case5" "Case6" "Case7"

$var.data.vector

F6 F6.1

Case1 0.00 90.00

Case2 -90.00 98.00

Case3 65.00 90.00

Case4 45.00 89.00

Case5 20.00 40.00

Case6 5.00 8.00

Case7 3.14 6.76

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

sym.obj

Examples

```
data(example3)
sym.data<-example3
sym.var(sym.data,4)
sym.var(sym.data,6)
```

---

sym.variance *Symbolic Variance*

Description

Compute the symbolic variance

Usage

```
sym.variance(sym.var, method = c("centers", "interval",
                     "billard", "histogram"), na.rm = FALSE, ...)
```

Arguments

sym.var	The symbolic variable.
method	The method to be use.
na.rm	As in R median function.
...	As in R median function.

Value

Return a real number.

Author(s)

Oldemar Rodriguez Rojas

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Rodriguez, O. (2000). Classification et Modeles Lineaires en Analyse des Donnees Symboliques. Ph.D. Thesis, Paris IX-Dauphine University.

Examples

```
data(example3)
sym.data<-example3
sym.variance(sym.var(sym.data,1))
sym.variance(sym.var(sym.data,2))
sym.variance(sym.var(sym.data,6))
sym.variance(sym.var(sym.data,6),method='interval')
sym.variance(sym.var(sym.data,6),method='billard')
sym.variance(sym.var(sym.data,3),method='histogram')
```

USCrime *Us crime classic data table*

Description

Us crime classic data table that can be used to generate symbolic data tables.

Usage

```
data(USCrime)
```

Source

http://archive.ics.uci.edu/ml/

References

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction. New York: Springer.

Examples

```
data(USCrime)
us.crime<-USCrime
dim(us.crime)
head(us.crime)
summary(us.crime)
names(us.crime)
nrow(us.crime)
result  <- classic.to.sym(us.crime, concept="state",
                          variables=c("NumInShelters", "NumImmig"),
                          variables.types=c("$H", "$H"))
result
```

uscrime_int *Us crime interval data table*

Description

Us crime classic data table genetated from uscrime data.

Usage

```
data(uscrime_int)
```

References

Rodriguez O. (2013). A generalization of Centre and Range method for fitting a linear regression model to symbolic interval data using Ridge Regression, Lasso and Elastic Net methods. The IFCS2013 conference of the International Federation of Classification Societies, Tilburg University Holland.

Examples

```
data(uscrime_int)
car.data<-uscrime_int
res.cm.lasso<-sym.glm(sym.data=car.data,response=102,method='cm',alpha=1,
                                        nfolds=10,grouped=TRUE)
plot(res.cm.lasso)
plot(res.cm.lasso$glmnet.fit, "norm", label=TRUE)
plot(res.cm.lasso$glmnet.fit, "lambda", label=TRUE)

pred.cm.lasso<-predictsym.glm(res.cm.lasso,response=102,car.data,method='cm')
RMSE.L(sym.var(car.data,102),pred.cm.lasso)
RMSE.U(sym.var(car.data,102),pred.cm.lasso)
R2.L(sym.var(car.data,102),pred.cm.lasso)
R2.U(sym.var(car.data,102),pred.cm.lasso)
deter.coefficient(sym.var(car.data,102),pred.cm.lasso)
```

VeterinaryData *Symbolic interval data example*

Description

Symbolic data matrix with all the variables of interval type

Usage

```
data(VeterinaryData)
```

Format

$I Height Height $I Weight Weight

1 $I 120.0 180.0 $I 222.2 354.0

2 $I 158.0 160.0 $I 322.0 355.0

3 $I 175.0 185.0 $I 117.2 152.0

4 $I 37.9 62.9 $I 22.2 35.0

5 $I 25.8 39.6 $I 15.0 36.2

6 $I 22.8 58.6 $I 15.0 51.8

7 $I 22.0 45.0 $I 0.8 11.0

8 $I 18.0 53.0 $I 0.4 2.5

9 $I 40.3 55.8 $I 2.1 4.5

10 $I 38.4 72.4 $I 2.5 6.1

References

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

Examples

```
data(VeterinaryData)
display.sym.table(VeterinaryData)
```

`write.sym.table` *Write Symbolic Data Table*

Description

This function write (save) a symbolic data table from a CSV data file.

Usage

```
write.sym.table(sym.data, file, sep, dec, row.names = NULL, col.names = NULL)
```

Arguments

`sym.data`	Symbolic data table
`file`	The name of the CSV file.
`sep`	As in R function read.table
`dec`	As in R function read.table
`row.names`	As in R function read.table
`col.names`	As in R function read.table

Value

Write in CSV file the symbolic data table.

Author(s)

Oldemar Rodriguez Rojas

References

Bock H-H. and Diday E. (eds.) (2000). Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data. Springer, Germany.

See Also

read.sym.table

Examples

```
data(example1)
write.sym.table(example1, file='temp4.csv', sep='|',dec='.', row.names=TRUE,
                col.names=TRUE)
ex1<-read.sym.table('temp4.csv', header=TRUE, sep='|',dec='.', row.names=1)
```

Index