Package 'SIN'

February 19, 2015

Version 0.6

Date 2013-12-23

Title A SINful Approach to Selection of Gaussian Graphical Markov Models

Author Mathias Drton <md5@uw.edu>.

Maintainer Mathias Drton <md5@uw.edu>

Depends R (>= 1.8.0)

Description This package provides routines to perform SIN model selection as described in Drton & Perlman (2004, 2008). The selected models are represented in the format of the 'ggm' package, which allows in particular parameter estimation in the selected model.

License GPL (>= 2)

URL http://www.r-project.org, http://www.stat.washington.edu/~md5

NeedsCompilation no

Repository CRAN

Date/Publication 2013-12-27 00:54:22

R topics documented:

`ambition` *Ambition and Attainment*

Description

In a study of ambition and attainment 7 variables are measured on 767 twelth grade males. The variables are:

X_1 : father's education,
X_2 : father's occupation,
X_3 : number of siblings,
X_4 : intelligence,
X_5 : grades,
X_6 : educational expectation,
X_7 : occupational aspiration.

Usage

```
data(ambition)
```

Format

A list providing a summary of the data. The list contains:

`ambition$means` : the mean vector,
`ambition$stddev` : the vector of standard deviations,

`ambition$corr`	:	the correlation matrix, and
`ambition$n`	:	the sample size.

Source

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See Ch. 10.4).

Examples

```
data(ambition)
ambition$means
ambition$stddev
ambition$corr
ambition$n
```

`anxietyanger` *Anxiety and Anger*

Description

Measurements on 684 students. The 4 variables are:

- Anxiety state,
- Anger state,
- Anxiety trait,
- Anger trait.

Usage

```
data(anxietyanger)
```

Format

A list providing a summary of the data. The list contains:

`anxietyanger$means`	:	the mean vector,
`anxietyanger$stddev`	:	the vector of standard deviations,
`anxietyanger$corr`	:	the correlation matrix, and
`anxietyanger$n`	:	the sample size.

Source

Cox, D.R. \& Wermuth, N. (1993) Linear Dependencies Represented by Chain Graphs. *Statistical Science* 8(3): 204-283. (See Table 1).

Examples

```
data(anxietyanger)
anxietyanger$means
anxietyanger$stddev
anxietyanger$corr
anxietyanger$n
```

`blauduncan` *Blau and Duncan's data on the American occupational structure*

Description

Sample of 20,700 subjects for which 5 variables are measured:

V	:	Father's educational attainment,
X	:	Father's occupational status,
U	:	educational attainment,
W	:	Status of first-job,
Y	:	Status of occupation in 1962.

Usage

```
data(blauduncan)
```

Format

A list providing a summary of the data. The list contains:

`blauduncan$means`	:	the mean vector,
`blauduncan$stddev`	:	the vector of standard deviations,
`blauduncan$corr`	:	the correlation matrix, and
`blauduncan$n`	:	the sample size.

Source

Blau, P. \& Duncan, O. (1967) *The American Occupational Structure*. New York: Wiley. (See p.13 and p.169).

References

Spirtes, P., Glymour, C. \& Scheines, R. (2000) *Causation, Prediction, and Search*. Cambridge, Massachusetts: The MIT Press. (See pp.105-108).

Examples

```
data(blauduncan)
blauduncan$means
blauduncan$stddev
```

```
blauduncan$corr
blauduncan$n
```

`bloodpressure` *Blood Pressure*

Description

In a study on relating blood pressure to age, weight, and stress at work, 6 measurements are taken on 98 subjects. The variables are:

a	:	age,
b	:	weight,
c	:	anger at work,
d	:	anxiety at work,
e	:	diastolic blood pressure,
f	:	systolic blood pressure.

Usage

```
data(bloodpressure)
```

Format

A list providing a summary of the data. The list contains:

`bloodpressure$means`	:	the mean vector,
`bloodpressure$stddev`	:	the vector of standard deviations,
`bloodpressure$corr`	:	the correlation matrix, and
`bloodpressure$n`	:	the sample size.

Source

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See p.343).

Examples

```
data(bloodpressure)
bloodpressure$means
bloodpressure$stddev
bloodpressure$corr
bloodpressure$n
```

`bodyfat` *Body Fat*

Description

Data consists of estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. The 15 variables are:

- Density determined from underwater weighing,
- Percent body fat,
- Age (years),
- Weight (lbs),
- Height (inches),
- Neck circumference (cm),
- Chest circumference (cm),
- Abdomen 2 circumference (cm),
- Hip circumference (cm),
- Thigh circumference (cm),
- Knee circumference (cm),
- Ankle circumference (cm),
- Biceps (extended) circumference (cm),
- Forearm circumference (cm),
- Wrist circumference (cm).

Usage

```
data(bodyfat)
```

Format

A list providing a summary of the data. The list contains:

`bodyfat$means`	:	the mean vector,
`bodyfat$stddev`	:	the vector of standard deviations,
`bodyfat$corr`	:	the correlation matrix, and
`bodyfat$n`	:	the sample size.

Source

The data set is available online at StatLib—Datasets Archive
http://lib.stat.cmu.edu/datasets/bodyfat .

Examples

```
data(bodyfat)
bodyfat$means
bodyfat$stddev
bodyfat$corr
bodyfat$n
```

`corkborings` *Cork Borings*

Description

Data from cork borings on 28 trees. The 4 variables are the weights of cork deposits in the four directions

N : North,
E : East,
S : South,
W : West.

Usage

```
data(corkborings)
```

Format

A list providing a summary of the data. The list contains:

`corkborings$means` : the mean vector,
`corkborings$stddev` : the vector of standard deviations,
`corkborings$corr` : the correlation matrix, and
`corkborings$n` : the sample size.

Source

Mardia, K.V., Kent, J.T. \& Bibby, J.M. (1979) *Multivariate analysis*. London: Academic Press. (See pp.11-12).

Examples

```
data(corkborings)
corkborings$means
corkborings$stddev
corkborings$corr
corkborings$n
```

fisherz *Fisher's z-transform*

Description

This function implements Fisher's z-transform, which maps a correlation $\rho \in (-1, 1)$ to

$$\frac{1}{2} \log \frac{1+\rho}{1-\rho}.$$

Usage

```
fisherz(corrs)
```

Arguments

corrs	a vector of correlations in (-1,1).

Value

The return value is a vector of z-transformed correlations.

References

Anderson, T.W. (2003) *An Introduction to Multivariate Statistical Analysis*, third edition. Hoboken, New Jersey: Wiley. (See p.133-134).

Examples

```
corrs <- c(-0.5,0,0.5)
fisherz(corrs)
```

fowlbones *Fowl Bones*

Description

Bone measurements are taken on 276 white leghorn fowl. The 6 variables are:

- skull length,
- skull breadth,
- humerous (wings),
- ulna (wings),
- femur (legs),
- tibia (legs).

Usage

```
data(fowlbones)
```

Format

A list providing a summary of the data. The list contains:

`fowlbones$means`	:	the mean vector,
`fowlbones$stddev`	:	the vector of standard deviations,
`fowlbones$corr`	:	the correlation matrix, and
`fowlbones$n`	:	the sample size.

Source

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See p.266).

Examples

```
data(fowlbones)
fowlbones$means
fowlbones$stddev
fowlbones$corr
fowlbones$n
```

`fretsheads` *Frets' Heads*

Description

Head measurements on the first an the second adult son in a sample of 25 families. The 4 variables are:

- head length (first son),
- head breadth (first son),
- head length (second son),
- head breadth (second son).

Usage

```
data(fretsheads)
```

Format

A list providing a summary of the data. The list contains:

`fretsheads$means`	:	the mean vector,
`fretsheads$stddev`	:	the vector of standard deviations,
`fretsheads$corr`	:	the correlation matrix, and
`fretsheads$n`	:	the sample size.

Source

Mardia, K.V., Kent, J.T. \& Bibby, J.M. (1979) *Multivariate analysis*. London: Academic Press. (See p.121).

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See p.265).

Examples

```
data(fretsheads)
fretsheads$means
fretsheads$stddev
fretsheads$corr
fretsheads$n
```

`getgraph` *Obtain graph from simultaneous p-values*

Description

This function converts a matrix of simultaneous p-values into a graph by comparing the p-values to a user-provided significance level.

Usage

```
getgraph(pvals, alpha, type="UG", blocks=NULL)
```

Arguments

`pvals`	a matrix of simultaneous p-values.
`alpha`	a significance level, i.e., alpha in (0,1).
`type`	a string specifying the type of graph that should be obtained from the p-value matrix. If type equals `"UG"` then an undirected graph is returned, if type equals `"DAG"` then an acyclic directed graph is returned, and if type equals `"BG"` then a bidirected graph is returned. If type equals the fourth possible choice `"CG"` then a chain graph is returned, in which case a list of integer vectors has to be provided as the input `blocks`.
`blocks`	a list of integer vectors specifying a family of subsets of the variables.

Value

The function returns an adjacency matrix A with $A_{ij} = 0$ if there is no edge between vertices (variables) i and j. The convention for edges is that $i - j$ if and only if $A_{ij} = A_{ji} = 1$, $i \to j$ if and only if $A_{ij} = 1$ and $A_{ji} = 0$, and $i \leftrightarrow j$ if and only if $A_{ij} = A_{ji} = 2$.

Examples

```
data(fowlbones)
pvals <- sinUG(fowlbones$corr,fowlbones$n)
alpha <- 0.2
## get undirected graph
getgraph(pvals, alpha, type="UG")
## forget that we used sinUG and get acyclic directed graph
getgraph(pvals, alpha, type="DAG")
## forget that we used sinUG and get bidirected graph
getgraph(pvals, alpha, type="BG")
## forget that we used sinUG and get chain graph
myblocks <- list(1:2,3:4,5:6)
getgraph(pvals, alpha, type="CG", blocks=myblocks)
```

`glucose` *Glucose*

Description

From an investigation of determinants of blood glucose control. Data on 39 diabetic patients. The 4 variables are:

- glycosylated hemoglobin GHb,
- knowledge about the illness,
- duration of the illness,
- fatalism (measure of patient's attitude to disease).

Usage

```
data(glucose)
```

Format

A list providing a summary of the data. The list contains:

`glucose$means`	:	the mean vector,
`glucose$stddev`	:	the vector of standard deviations,
`glucose$corr`	:	the correlation matrix, and
`glucose$n`	:	the sample size.

Source

Cox, D.R. \& Wermuth, N. (1993) Linear Dependencies Represented by Chain Graphs. *Statistical Science* 8(3): 204-283. (See Table 7).

Examples

```
data(glucose)
glucose$means
glucose$stddev
glucose$corr
glucose$n
```

`grades` *School Grades*

Description

The data consists of the test results of 220 boys in the 6 school subjects:

- Gaelic,
- English,
- history,
- arithmetic,
- algebra,
- geometry.

Usage

```
data(grades)
```

Format

A list providing a summary of the data. The list contains:

`grades$means`	:	the mean vector,
`grades$stddev`	:	the vector of standard deviations,
`grades$corr`	:	the correlation matrix, and
`grades$n`	:	the sample size.

Source

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See p.266).

Examples

```
data(grades)
grades$means
grades$stddev
grades$corr
grades$n
```

hiv *HIV*

Description

The data consists of 6 blood measurements taken on 107 babies:

- immunoglobin G,
- immunoglobin A,
- lymphocyte B,
- platelet count,
- lymphocyte T4,
- T4/T8 lymphocyte ratio.

Usage

```
data(hiv)
```

Format

A list providing a summary of the data. The list contains:

`hiv$means`	:	the mean vector,
`hiv$stddev`	:	the vector of standard deviations,
`hiv$corr`	:	the correlation matrix, and
`hiv$n`	:	the sample size.

Source

Roverato, A. \& Whittaker, J. (1996) A hyper normal prior distribution for approximate Bayes factor calculations on non-decomposable graphical Gaussian models, unpublished manuscript.

Examples

```
data(hiv)
hiv$means
hiv$stddev
hiv$corr
hiv$n
```

holm *Holm's step-down p-values*

Description

This function takes a matrix of simultaneous p-values from Sidak's inequality and adjusts the p-values according to Holm's step-down procedure.

Usage

```
holm(pvals)
```

Arguments

pvals	a matrix of simultaneous p-values from Sidak's inequality.

Value

A matrix of simultaneous p-values from Holm's step-down procedure with NA on the diagonal.

References

Holm, S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6: 65-70.

Examples

```
data(mathmarks)
sinUG(mathmarks$corr,mathmarks$n)
holm(sinUG(mathmarks$corr,mathmarks$n))
```

---

is.blocks *Check variable blocking*

Description

This function checks whether a list of vectors defines a valid decomposition of the set of variables into blocks.

Usage

```
is.blocks(blocks, p)
```

Arguments

blocks	a list of integer vectors representing a family of subsets of the variable set.
p	the number of variables.

Value

The function returns TRUE if the block structure is such that each variable is represented in exactly one of the blocks. Otherwise, FALSE is returned.

Examples

```
p <- 6
blocks <- list(1:3,6,5:4)
is.blocks(blocks, p)
blocks <- list(1:3,7,5:4)
is.blocks(blocks, p)
blocks <- list(1:2,6,5:4)
is.blocks(blocks, p)
```

---

mathmarks *Mathematics marks*

Description

Examination marks of 88 students in the five subjects:

- Mechanics,
- Vectors,
- Algebra,
- Analysis,
- Statistics.

Usage

```
data(mathmarks)
```

Format

A list providing a summary of the data. The list contains:

`mathmarks$means`	:	the mean vector,
`mathmarks$stddev`	:	the vector of standard deviations,
`mathmarks$corr`	:	the correlation matrix, and
`mathmarks$n`	:	the sample size.

Details

Mechanics and Vectors were closed book examinations. Algebra, Analysis and Statistics were open book examinations.

Source

Mardia, K.V., Kent, J.T. \& Bibby, J.M. (1979) *Multivariate analysis*. London: Academic Press. (See pp.3-4).

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See Ch. 1.1).

Examples

```
data(mathmarks)
mathmarks$means
mathmarks$stddev
mathmarks$corr
mathmarks$n
```

`moth` *Noctuid Moth Trappings*

Description

The data are of sample size 72 and contain 6 variables:

min minimum night temperature,

max previous day's maximum temperature,

wind average wind speed during night,

rain amount of rain during night,

cloud percentage of starlight obscured by clouds,

moth number of moths caught in a light trap in one night.

Usage

```
data(moth)
```

Format

A list providing a summary of the data. The list contains:

`moth$means`	:	the mean vector,
`moth$stddev`	:	the vector of standard deviations,
`moth$corr`	:	the correlation matrix, and
`moth$n`	:	the sample size.

Source

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See Ch. 10.3).

Examples

```
data(moth)
moth$means
moth$stddev
moth$corr
moth$n
```

plotBGpvalues *Plot simultaneous p-values for bidirected graphs*

Description

This function plots the simultaneous p-values for SIN model selection for bidirected graphs.

Usage

```
plotBGpvalues(pvals, legend=TRUE, legendpos=NULL)
```

Arguments

pvals	a matrix of p-values.
legend	a Boolean.
legendpos	a numerical vector of length 2.

Value

A plot illustrating the simultaneous p-values. If legend==TRUE, then a legend giving information on the variable labelling is drawn. Optionally, the position of the legend can be specified by legendpos.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

sinBG

Examples

```
data(stressful)
pvals <- holm(sinBG(stressful$corr,stressful$n))
## Not run: plotBGpvalues(pvals)
## Not run: plotBGpvalues(pvals, legend=FALSE)
## Not run: plotBGpvalues(pvals, legendpos=c(5,0.5))
## Not run: plotBGpvalues(pvals, legend=TRUE, legendpos=c(5,0.5))
```

`plotCGpvalues` *Plot simultaneous p-values for chain graphs*

Description

This function plots the simultaneous p-values for SIN model selection for chain graphs.

Usage

```
plotCGpvalues(blocks, pvals, legend=TRUE, legendpos=NULL)
```

Arguments

`blocks`	a list of integer vectors with entries amongst $1, \ldots, p$ where p is the number of variables. Each one of the integer vectors specifies a set of variables that form a block in the chain graph. Furthermore, a partial ordering of the variables is specified by the convention that variables in one block are ordered smaller than variables in a block succeeding in the list `blocks`.
`pvals`	a matrix of p-values.
`legend`	a Boolean.
`legendpos`	a numerical vector of length 2.

Value

A plot illustrating the simultaneous p-values. If `legend==TRUE`, then a legend giving information on the variable labelling is drawn. Optionally, the position of the legend can be specified by `legendpos`.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

`sinCG`

Examples

```
data(fowlbones)
blocks <- list(1:2,3:4,5:6)
pvals <- holm(sinCG(blocks,fowlbones$corr,fowlbones$n, type="AMP"))
## Not run: plotCGpvalues(blocks, pvals)
## Not run: plotCGpvalues(blocks, pvals, legend=FALSE)
```

```
## Not run: plotCGpvalues(blocks, pvals, legendpos=c(7,0.5))
## Not run: plotCGpvalues(blocks, pvals, legend=TRUE, legendpos=c(7,0.5))
```

`plotDAGpvalues` *Plot simultaneous p-values for acyclic directed graphs*

Description

This function plots the simultaneous p-values for SIN model selection for acyclic directed graphs (DAGs).

Usage

```
plotDAGpvalues(pvals, legend=TRUE, legendpos=NULL)
```

Arguments

`pvals`	a matrix of p-values.
`legend`	a Boolean.
`legendpos`	a numerical vector of length 2.

Value

A plot illustrating the simultaneous p-values. If `legend==TRUE`, then a legend giving information on the variable labelling is drawn. Optionally, the position of the legend can be specified by `legendpos`.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

`sinDAG`

Examples

```
data(fowlbones)
p <- dim(fowlbones$corr)[1]
pvals <- holm(sinDAG(list(1:p),fowlbones$corr,fowlbones$n))
## Not run: plotDAGpvalues(pvals)
## Not run: plotDAGpvalues(pvals, legend=FALSE)
## Not run: plotDAGpvalues(pvals, legendpos=c(7,0.5))
## Not run: plotDAGpvalues(pvals, legend=TRUE, legendpos=c(7,0.5))
```

`plotUGpvalues` *Plot simultaneous p-values for undirected graphs*

Description

This function plots the simultaneous p-values for SIN model selection for undirected graphs.

Usage

```
plotUGpvalues(pvals, legend=TRUE, legendpos=NULL)
```

Arguments

`pvals`	a matrix of p-values.
`legend`	a Boolean.
`legendpos`	a numerical vector of length 2.

Value

A plot illustrating the simultaneous p-values. If `legend==TRUE`, then a legend giving information on the variable labelling is drawn. Optionally, the position of the legend can be specified by `legendpos`.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

`sinUG`

Examples

```
data(fowlbones)
pvals <- holm(sinUG(fowlbones$corr,fowlbones$n))
## Not run: plotUGpvalues(pvals)
## Not run: plotUGpvalues(pvals, legend=FALSE)
## Not run: plotUGpvalues(pvals, legendpos=c(7,0.5))
## Not run: plotUGpvalues(pvals, legend=TRUE, legendpos=c(7,0.5))
```

pubprod *Publishing productivity*

Description

Sample of 162 subjects for which 7 variables are measured:

ability	:	score of the subject's ability,
GPQ	:	measure of the quality of the graduate program attended ,
preprod	:	preliminary measure of productivity,
QFJ	:	quality of the first-job,
sex	:	sex,
cites	:	citation rate,
pubs	:	publication rate.

Usage

```
data(pubprod)
```

Format

A list providing a summary of the data. The list contains:

`pubprod$means`	:	the mean vector,
`pubprod$stddev`	:	the vector of standard deviations,
`pubprod$corr`	:	the correlation matrix, and
`pubprod$n`	:	the sample size.

Source

Spirtes, P., Glymour, C. \& Scheines, R. (2000) *Causation, Prediction, and Search*. Cambridge, Massachusetts: The MIT Press. (See Example 5.8.1).

Examples

```
data(pubprod)
pubprod$means
pubprod$stddev
pubprod$corr
pubprod$n
```

sdcor2cov *Covariance matrix*

Description

This function takes a vector of standard deviations and a correlation matrix as input and computes the covariance matrix.

Usage

```
sdcor2cov(stddev, corr)
```

Arguments

stddev	a vector of standard deviations.
corr	a correlation matrix.

Value

The function returns the covariance matrix corresponding to the input information.

Examples

```
data(sur)
sdcor2cov(sur$stddev, sur$corr)
```

---

sinBG *SIN for bidirected graphs*

Description

This function computes the matrix of simultaneous p-values for SIN model selection for bidirected (or covariance) graphs.

Usage

```
sinBG(S,n,holm=TRUE)
```

Arguments

S	a covariance or correlation matrix.
n	the sample size,
holm	Boolean variable indicating whether Holm's p-value adjustment should be used (holm=TRUE) or not (holm=FALSE).

Value

A matrix of simultaneous p-values with NA on the diagonal.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

plotBGpvalues

Examples

```
data(stressful)
sinBG(stressful$corr,stressful$n)
sinBG(stressful$corr,stressful$n,holm=FALSE)
```

---

sinCG *SIN for chain graphs*

Description

This function computes the matrix of simultaneous p-values for SIN model selection for chain graphs. SIN assumes that a dependence chain or blocking of the variables is available.

Usage

```
sinCG(blocks, S, n, type="AMP", holm=TRUE)
```

Arguments

blocks	a list of integer vectors with entries amongst $1, \ldots, p$ where p is the number of variables. Each one of the integer vectors specifies a set of variables that form a block in the chain graph. Furthermore, a partial ordering of the variables is specified by the convention that variables in one block are ordered smaller than variables in a block succeeding in the list blocks.
S	a covariance or correlation matrix.
n	the sample size.
type	a string equal to either "AMP" or "LWF" that determines whether the chain graph is interpreted with respect to the Andersson, Madigan, Perlman Markov property or the Lauritzen, Wermuth, Frydenberg Markov property.
holm	Boolean variable indicating whether Holm's p-value adjustment should be used (holm=TRUE) or not (holm=FALSE).

Value

A matrix of simultaneous p-values with NA on the diagonal.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

Andersson, S.A., Madigan, D. \& Perlman, M.D. (2001) Alternative Markov Properties for Chain Graphs. *Scandinavian Journal of Statistics* 28(1): 33-85.

Lauritzen, S. (1996) *Graphical Models*. Oxford University Press: Oxford.

See Also

plotCGpvalues

Examples

```
data(fowlbones)
p <- dim(fowlbones$corr)[1]
blocks <- list(1:2,3:4,5:6)
sinCG(blocks,fowlbones$corr,fowlbones$n, type="AMP")
sinCG(blocks,fowlbones$corr,fowlbones$n, type="LWF")
sinCG(blocks,fowlbones$corr,fowlbones$n, type="AMP", holm=FALSE)
sinCG(blocks,fowlbones$corr,fowlbones$n, type="LWF", holm=FALSE)
```

---

sinDAG *SIN for acyclic directed graphs*

Description

This function computes the matrix of simultaneous p-values for SIN model selection for acyclic directed graphs (DAGs). SIN assumes that a total ordering of the variables is available (e.g., as a time-order).

Usage

```
sinDAG(order,S,n,holm=TRUE)
```

Arguments

order	a list of pairwise different integers amongst $1, \ldots, p$ where p is the number of variables. This list specifies a total ordering of the variables by the convention that a list element is smaller than succeeding list elements.
S	a covariance or correlation matrix.
n	the sample size.
holm	Boolean variable indicating whether Holm's p-value adjustment should be used (holm=TRUE) or not (holm=FALSE).

Value

A matrix of simultaneous p-values with NA on the diagonal.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

plotDAGpvalues

Examples

```
data(fowlbones)
p <- dim(fowlbones$corr)[1]
sinDAG(list(1:p),fowlbones$corr,fowlbones$n)
sinDAG(list(1:p),fowlbones$corr,fowlbones$n,holm=FALSE)
sinDAG(list(3,2,1,4,5,6),fowlbones$corr,fowlbones$n)
```

---

sinUG *SIN for undirected graphs*

Description

This function computes the matrix of simultaneous p-values for SIN model selection for undirected (or concentration) graphs.

Usage

```
sinUG(S,n,holm=TRUE)
```

Arguments

S	a covariance or correlation matrix.
n	the sample size.
holm	Boolean variable indicating whether Holm's p-value adjustment should be used (holm=TRUE) or not (holm=FALSE).

Value

A matrix of simultaneous p-values with NA on the diagonal.

References

Drton, M. \& Perlman, M.D. (2004) Model Selection for Gaussian Concentration Graphs. *Biometrika* 91(3): 591-602.

Drton, M. \& Perlman, M.D. (2008) A SINful Approach to Gaussian Graphical Model Selection. *J. Statist. Plann. Inference* 138(4): 1179-1200.

See Also

plotUGpvalues

Examples

```
data(fowlbones)
sinUG(fowlbones$corr,fowlbones$n)
sinUG(fowlbones$corr,fowlbones$n, holm=FALSE)
```

`socstatus` *Social Status and Participation*

Description

In a study on the relationship between social status and participation, 6 variables are measured for a sample of 530 women:

X_1 : income,
X_2 : occupation,
X_3 : education,
Y_1 : church attendance,
Y_2 : memberships,
Y_3 : friends' seen.

Usage

```
data(socstatus)
```

Format

A list providing a summary of the data. The list contains:

`socstatus$means` : the mean vector,
`socstatus$stddev` : the vector of standard deviations,
`socstatus$corr` : the correlation matrix, and
`socstatus$n` : the sample size.

Source

Whittaker, J. (1990) *Graphical models in applied multivariate statistics*. Chichester: Wiley. (See p.327).

Examples

```
data(socstatus)
socstatus$means
socstatus$stddev
socstatus$corr
socstatus$n
```

`stressful` *Stressful Events*

Description

Results of questionnaires given to 72 students. The 4 variables are:

- cognitive avoidance,
- vigilance,
- blunting,
- monitoring.

Usage

```
data(stressful)
```

Format

A list providing a summary of the data. The list contains:

`stressful$means`	:	the mean vector,
`stressful$stddev`	:	the vector of standard deviations,
`stressful$corr`	:	the correlation matrix, and
`stressful$n`	:	the sample size.

Source

Cox, D.R. \& Wermuth, N. (1993) Linear Dependencies Represented by Chain Graphs. *Statistical Science* 8(3): 204-283. (See Table 3).

Examples

```
data(stressful)
stressful$means
stressful$stddev
```

```
stressful$corr
stressful$n
```

---

sur *Simulated Data from Seemingly Unrelated Regressions*

Description

The data are simulated from a bivariate seemingly unrelated regressions model with two response variables (Y_1, Y_2) and two covariates (X_1, X_2). The sample is only of size 8.

Usage

```
data(sur)
```

Format

A list providing a summary of the data. The list contains:

`sur$means`	:	the mean vector,
`sur$stddev`	:	the vector of standard deviations,
`sur$corr`	:	the correlation matrix, and
`sur$n`	:	the sample size.

Details

The data were used by Drton \& Richardson (2004) to illustrate the possibility of a multimodal likelihood in seemingly unrelated regressions.

Source

Drton, M. \& Richardson, T.S. (2004) Multimodality of the likelihood in the bivariate seemingly unrelated regressions model. *Biometrika* 91(2): 383-392.

Examples

```
data(sur)
sur$means
sur$stddev
sur$corr
sur$n
```

---

university1992 *Druzdzel and Glymour's University Data 1992*

Description

Data from a study by the *U.S. News and World Report* magazine with the purpose of college ranking. Druzdzel and Glymour choose 8 variables and after removing some universities give a correlation matrix for 170 universities. The 8 variables are:

apgra average percentage of graduation, i.e., the fraction of the total number of entering students who make it through the graduation,

rejr rejection rate, i.e., the fraction of the applicants who are given an admission offer,

tstsc average standardized test scores of the incoming students,

top10 class standing of the incoming freshman, which is a fraction of the incoming freshmen who were in top 10% of their high school graduating class,

pacc percentage of those students who accept the university's offer from among those who are offered admission,

spend total educational and general expenses per student, which is the sum spent on the instruction, student services, and academic support, including libraries and computing services,

strat student-faculty ratio,

salar average facutly salary.

Usage

```
data(university1992)
```

Format

A list providing a summary of the data. The list contains:

`university1992$means`	:	the mean vector,
`university1992$stddev`	:	the vector of standard deviations,
`university1992$corr`	:	the correlation matrix, and
`university1992$n`	:	the sample size.

Source

Druzdzel, M.J. \& Glymour, C. (1999) In *Computation, Causation, and Discovery*. (Ed. Glymour, C. \& Cooper, G.F.). Cambridge, Massachusetts: The MIT Press. (See Ch. 19).

Examples

```
data(university1992)
university1992$means
university1992$stddev
university1992$corr
university1992$n
```

`university1993` *Druzdzel and Glymour's University Data 1993*

Description

Data from a study by the *U.S. News and World Report* magazine with the purpose of college ranking. Druzdzel and Glymour choose 8 variables and after removing some universities give a correlation matrix for 159 universities. The 8 variables are:

apgra average percentage of graduation, i.e., the fraction of the total number of entering students who make it through the graduation,

rejr rejection rate, i.e., the fraction of the applicants who are given an admission offer,

tstsc average standardized test scores of the incoming students,

top10 class standing of the incoming freshman, which is a fraction of the incoming freshmen who were in top 10% of their high school graduating class,

pacc percentage of those students who accept the university's offer from among those who are offered admission,

spend total educational and general expenses per student, which is the sum spent on the instruction, student services, and academic support, including libraries and computing services,

strat student-faculty ratio,

salar average facutly salary.

Usage

```
data(university1993)
```

Format

A list providing a summary of the data. The list contains:

`university1993$means`	:	the mean vector,
`university1993$stddev`	:	the vector of standard deviations,
`university1993$corr`	:	the correlation matrix, and
`university1993$n`	:	the sample size.

Source

Druzdzel, M.J. \& Glymour, C. (1999) In *Computation, Causation, and Discovery*. (Ed. Glymour, C. \& Cooper, G.F.). Cambridge, Massachusetts: The MIT Press. (See Ch. 19).

Examples

```
data(university1993)
university1993$means
university1993$stddev
university1993$corr
```

```
university1993$n
```

Index