Package ‘MAINT.Data’

Type Package
Title Model and Analyse Interval Data
Version 0.3
Date 2014-10-24
Author Pedro Duarte Silva <psilva@porto.ucp.pt> Paula Brito <mpbrito.fep.up.pt>
Maintainer Pedro Duarte Silva <psilva@porto.ucp.pt>
Description MAINT.Data implements methodologies for modelling interval data, considering appropriate parameterizations of the variance-covariance matrix that take into account the intrinsic nature of interval data, and lead to five different possible configuration structures.
License GPL-2
LazyLoad yes
LazyData yes
Depends R (>= 2.8.0)
Imports MASS, methods, miscTools
Repository CRAN
NeedsCompilation no
Date/Publication 2014-10-25 01:27:10

R topics documented:

MAINT.Data-package .......................... 2
BestModel-methods .......................... 4
ChinaTemp ................................. 5
ConfTests-class ......................... 5
DACrossVal .............................. 6
IData ................................. 8
IData-class .............................. 9
IdtClMANOVA-class .................... 10
IdtE-class .............................. 11
IdtHetNMANOVA-class ................. 12
Idtlda-class ......................... 13
Description

MAINT.DATA implements methodologies for modelling Interval Data, considering five different possible configurations structures for the variance-covariance matrix. It introduces a data class for representing interval data and includes functions and methods for parametric modelling and analysing of interval data. It performs maximum likelihood estimation and statistical tests as well as (M)ANOVA and Linear and Quadratic Discriminant Analysis for all considered configurations.

Details

In the classical model of multivariate data analysis, data is represented in a data-array where n “individuals” (usually in rows) take exactly one value for each variable (usually in columns). Symbolic Data Analysis (see, e.g., Noirhomme-Fraiture and Brito (2011)) provides a framework where new variable types allow to take directly into account variability and/or uncertainty associated to each single “individual”, by allowing multiple, possibly weighted, values for each variable. New variable types - interval, categorical multi-valued and modal variables - have been introduced. We focus on the analysis of interval data, i.e., where elements are described by variables whose values are intervals. Parametric inference methodologies based on probabilistic models for interval variables are developed in Brito and Duarte Silva (2012) where each interval is represented by its midpoint and log-range, for which Normal and Skew-Normal distributions are assumed. The intrinsic nature of the interval variables leads to special structures of the variance-covariance matrix, which are represented by five different possible configurations.

MAINT.DATA implements the proposed methodologies in R, introducing a data class for representing interval data; it includes functions for modelling and analysing interval data, in particular maximum likelihood estimation and statistical tests for the different considered configurations. Methods for (M)ANOVA and Linear and Quadratic Discriminant Analysis of this data class are also provided.

Package: MAINT.Data
Type: Package
Version: 0.3
Date: 2014-10-24
License: GPL-2
LazyLoad: yes
LazyData: yes

Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>
Maintainer: Pedro Duarte Silva <psilva@porto.ucp.pt>

References


Examples
# Create an Interval-Data object containing the intervals for 899 observations
# on the temperatures by quarter in 60 Chinese meteorological stations.

ChinaT <- IData(ChinaTemp[1:8], VarNames=c("T1","T2","T3","T4"))

# Display the first and last observations
head(ChinaT)
tail(ChinaT)

# Print summary statistics
summary(ChinaT)

# Create a new data set considering only the Winter (1st and 4th) quarter intervals

ChinaWT <- ChinaT[,c(1,4)]

# Estimate parameters by maximum likelihood, assuming
# the classical (unrestricted) covariance configuration C1
ChinaWTE.C1 <- mle(ChinaWT, Config=1)
cat("Winter temperatures of China -- maximum likelihood estimation results:\n")
print(ChinaWTE.C1)
cat("Standard Errors of Estimators:\n") ; print(stdErr(ChinaWTE.C1))

# Estimate parameters by maximum likelihood,
# assuming that one of the C3, C4 or C5 restricted covariance configurations holds
ChinaWTE.C345 <- mle(ChinaWT, Config=3:5)
BestModel-methods

Methods for Function BestModel in Package ‘MAINT.Data’

Description

Selects the best model according to the chosen selection criterion (currently, AIC or BIC)

Usage

BestModel(Idt,SelCrit=c("IdtCrt","AIC","BIC"))

Arguments

Idt An IData object representing NObs interval-valued entities.
SelCrit The model selection criterion

Value

An integer with the index of the model chosen by the selection criterion
**Description**

This data set consists of the intervals of observed temperatures (Celsius scale) in each of the four quarters, Q_1 to Q_4, of the years 1974 to 1988 in 60 Chinese meteorological stations; one outlier observation (YinChuan_1982) has been discarded. The 60 stations belong to different regions in China, which therefore define a partition of the 899 stations-year combinations.

**Usage**

```r
data(ChinaTemp)
```

**Format**

A data frame containing 899 observations on 9 variables, the first eight with the lower and upper bounds of the temperatures by quarter in the 899 stations-year combinations, the last one a factor indicating the geographical region of each station.

**ConfTests-class**

Class "Configuration Tests"

**Description**

ConfTests contains a list of the results of statistical likelihood-ratio tests that evaluate the goodness-of-fit of variance-covariance matrix configurations against more general ones.

**Slots**

- **TestRes**: List of test results; each element is an object of type LRTest, with the following components: value of the Qui-Square statistics, degrees of freedom of the Qui-Square statistics, p-value of the Qui-Square statistics value, logarithm of the Likelihood function under the null hypothesis, logarithm of the Likelihood function under the alternative hypothesis.
- **RestModels**: An acronym describing the restricted model (corresponding to the null hypothesis).
- **FullModels**: An acronym describing the full model (corresponding to the alternative hypothesis).

**Methods**

- **show** signature(object = "ConfTests"): show S4 method for the ConfTests-class.

**Author(s)**

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>
References


See Also

mle, IData LRTest

DACrossVal

Cross Validation for Discriminant Analysis Classification Rules

Description

‘DACrossVal’ evaluates the performance of a Discriminant Analysis training sample algorithm by k-fold Cross-Validation.

Usage

DACrossVal(data, grouping, TrainAlg, EvalAlg=EvalClrule, Strfolds=TRUE, kfold=10, CVrep=20, prior="proportions", ...)

Arguments

data Matrix, data frame or Interval Data object of observations.

grouping Factor specifying the class for each observation.

TrainAlg A function with the training algorithm. It should return an object that can be used as input to the argument of ‘EvalAlg’.

EvalAlg A function with the evaluation algorithm. By default set to ‘EvalClrule’ which returns a list with components “err” (estimates of error rates by class) and “Nk” (number of out-sample observations by class). This default can be used for all ‘TrainAlg’ arguments that return an object with a predict method returning a list with a ‘class’ component (a factor) containing the classification results.

Strfolds Boolean flag indicating if the folds should be stratified according to the original class proportions (default), or randomly generated from the whole training sample, ignoring class membership.

kfold Number of training sample folds to be created in each replication.

CVrep Number of replications to be performed.

prior The prior probabilities of class membership. If unspecified, the class proportions for the training set are used. If present, the probabilities should be specified in the order of the factor levels.

... Further arguments to be passed to ‘TrainAlg’ and ‘EvalAlg’.
Value

A three dimensional array with the number of tested observations, and estimated classification errors for each combination of fold and replication tried. The array dimensions are defined as follows:

The first dimension runs through the different fold-replication combinations.
The second dimension represents the classes.
The third dimension has two named levels representing respectively the number of observations tested ("Nk"), and the estimated classification errors ("Clerr").

Author(s)

A. Pedro Duarte Silva

See Also

lda, qda, IData

Examples

# Compare performance of linear and quadratic discriminant analysis with
# Configurations C1 and c4 on the ChinaT data set by 5-fold cross-validation
# replicated twice

# Create an Interval-Data object containing the intervals for 899 observations
# on the temperatures by quarter in 60 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8])

# Classical (configuration 1) Linear Discriminant Analysis
CVldaC1 <- DACrossVal(ChinaT,ChinaTemp$GeoReg,TrainAlg=lda,Config=1,kfold=5,CVrep=2)
summary(CVldaC1[,,"Clerr"])
gberrors <-
apply(CVldaC1[,,"Nk"]*CVldaC1[,,"Clerr"],1,sum)/apply(CVldaC1[,,"Nk"],1,sum)
cat("Average global classification error =",mean(gberrors),"\n")

# Linear Discriminant Analysis with configuration 4
CVldaC4 <- DACrossVal(ChinaT,ChinaTemp$GeoReg,TrainAlg=lda,Config=4,kfold=5,CVrep=2)
summary(CVldaC4[,,"Clerr"])
gberrors <-
apply(CVldaC4[,,"Nk"]*CVldaC4[,,"Clerr"],1,sum)/apply(CVldaC4[,,"Nk"],1,sum)
cat("Average global classification error =",mean(gberrors),"\n")

# Classical (configuration 1) Quadratic Discriminant Analysis
CVqdaC1 <- DACrossVal(ChinaT,ChinaTemp$GeoReg,TrainAlg=qda,Config=1,kfold=5,CVrep=2)
summary(CVqdaC1[,,"Clerr"])
gberrors <-
apply(CVqdaC1[,,"Nk"]*CVqdaC1[,,"Clerr"],1,sum)/apply(CVqdaC1[,,"Nk"],1,sum)
cat("Average global classification error =",mean(gberrors),"\n")
# Quadratic Discriminant Analysis with configuration 4

```r
CVqdaC4 <- DAxCrossVal(ChinaT,ChinaTemp$GeoReg,TrainAlg=qda,Config=4,kfold=5,CVErep=2)
summary(CVqdaC4[,"Cllerr"])
glerrors <-
apply(CVqdaC4[,"Nk"]*CVqdaC4[,"Cllerr"],1,sum)/apply(CVqdaC4[,"Nk"],1,sum)
cat("Average global classification error =",mean(glerrors),"\n")
```

---

### IData

**Interval Data objects**

#### Description

`IData` creates `IData` objects from data frames of interval bounds or MidPoint/LogRange values of the interval-valued observations.

#### Usage

```r
IData(Data,
Seq = c("LbUb_VarbyVar", "MidPLogR_VarbyVar", "AllLb_AllUb",
"AllMidP_AllLogR"), VarNames=NULL, ObsNames=row.names(Data))
```

#### Arguments

- **Data**: a data frame of interval bounds or MidPoint/LogRange values.
- **Seq**: the format of ‘Data’ data frame. Available options are:
  - “LbUb_VarbyVar”: lower bounds followed by upper bounds, variable by variable.
  - “MidPLogR_VarbyVar”: MidPoints followed by LogRanges, variable by variable.
  - “AllLb_AllUb”: all lower bounds followed by all upper bounds, in the same variable order.
  - “AllMidP_AllLogR”: all MidPoints followed all LogRanges, in the same variable order.
- **VarNames**: An optional vector of names to be assigned to the Interval-Valued Variables.
- **ObsNames**: An optional vector of names assigned to the individual observations.

#### Details

Objects of type ‘IData’, describe a data set of ‘NObs’ observations on ‘NIVar’ Interval-Valued variables. This function creates an interval-data object from a data-frame with either the lower and upper bounds of the observed intervals or by their midpoints and log-ranges.
Examples

# Create an Interval-Data object containing the intervals for 899 observations # on the temperatures by quarter in 60 Chinese meteorological stations.

ChinaT <- IData(ChinaTemp[1:8], VarNames=c("T1","T2","T3","T4"))

cat("Summary of the ChinaT IData object:

print(summary(ChinaT))
cat("ChinaT first ant last observations:

print(head(ChinaT,n=3))
cat("...

print(tail(ChinaT,n=3))

Description

A data-array of interval-valued data is an array where each of the NObs rows, corresponding to each entity under analysis, contains the observed intervals of the NIVar descriptive variables.

Slots

MidP: A data-frame of the midpoints of the observed intervals
LogR: A data-frame of the logarithms of the ranges of the observed intervals
ObsNames: An optional vector of names assigned to the individual observations.
VarNames: An optional vector of names to be assigned to the Interval-Valued Variables.
NObs: Number of entities under analysis (cases)
NIVar: Number of interval variables

Methods

head signature(x = "IData"): head S4 method for the IData-class
show signature(object = "IData"): show S4 method for the IData-class
nrow signature(x = "IData"): returns the number of entities
ncol signature(x = "IData"): returns the number of Interval Variables
tail signature(x = "IData"): tail S4 method for the IData-class
mle signature(x = "IData"): Maximum likelihood estimation
MANOVA signature(x = "IData"): MANOVA tests on the interval-valued data
lda signature(x = "IData"): performs Linear Discriminant Analysis using the estimated model parameters
qda signature(x = "IData"): performs Quadratic Discriminant Analysis using the estimated model parameters
Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References


See Also
ldata, mle, MANOVA, ldd, qda

idtcmANOVA-class

Description
IdtClMANOVA contains the results of classical MANOVA tests applied to the chosen variables.

Slots
- NIVar: Number of interval variables
- grouping: Factor indicating the group to which the observation belongs to
- H0res: Model estimates under the null hypothesis
- H1res: Model estimates under the alternative hypothesis
- QuiSq: Value of the Qui-Square statistics corresponding to the performed test
- df: Degrees of freedom of the Qui-Square statistics
- pvalue: p-value of the Qui-Square statistics value, obtained from the Qui-Square distribution with df degrees of freedom
- H0logLik: Logarithm of the Likelihood function under the null hypothesis
- H1logLik: Logarithm of the Likelihood function under the alternative hypothesis

Extends
Class "ldtMANOVA", directly. Class "LRTest", by class "ldtMANOVA", distance 2.

Methods
lda signature(x = "IdtClMANOVA"): lda performs Linear Discriminat Analysis using the estimated model parameters
IdtE-class

Author(s)

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References


See Also

MANOVA, lda, IData

IdtE-class

Class "IdtE"

Description

IdtE contains the results of model estimation.

Slots

ModelNames: The model acronym, indicating the model type, currently N for Normal and the configuration (C1 to C5)
ModelType: Indicates the model; currently only Gaussian (Normal) distributions are implemented
ModelConfig: Configuration of the variance-covariance matrix: a vector of integers in the range 1 to 5
NIVar: Number of interval variables
SelCrit: The model selection criterion; currently, AIC and BIC are implemented
logLik: The logarithms of the likelihood function for the different cases
AICs: Value of the AIC criterion
BICs: Value of the BIC criterion
BestModel: The index of best model according to the chosen selection criterion

Methods

BestModel signature(Idt = "IdtE"): Selects the best model according to the chosen selection criterion (currently, AIC or BIC)
show signature(object = "IdtE"): show S4 method for the IdtE-class
testMod signature(Idt = "IdtE"): Performs statistical likelihood-ratio tests that evaluate the goodness-of-fit of a nested model against a more general one.
Author(s)

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References


See Also

mle, MANOVA, IData

IdtHetNMANOVA-class

Class "IdtHetNMANOVA"

Description

IdtHetNMANOVA contains the results of MANOVA tests on the interval-valued data, for the Het-erocedastic Normal model.

Slots

NIVar: Number of interval variables
grouping: Factor indicating the group to which each observation belongs to
H0res: Model estimates under the null hypothesis
H1res: Model estimates under the alternative hypothesis
Quisq: Value of the Qui-Square statistics corresponding to the performed test
df: Degrees of freedom of the Qui-Square statistics
pvalue: p-value of the Qui-Square statistics value, obtained from the Qui-Square distribution with df degrees of freedom
H0logLik: Logarithm of the Likelihood function under the null hypothesis
H1logLik: Logarithm of the Likelihood function under the alternative hypothesis

Extends

Class "IdtMANOVA", directly. Class "LRTest", by class "IdtMANOVA", distance 2.

Methods

qda signature(x = "IdtHetNMANOVA"): performs Quadratic Discriminant Analysis using the variance-covariance matrices estimated under Gaussian heterocedastic models.
Idtlda-class

Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References

See Also
MANOVA, qda, IData

Idtlda-class

Class "Idtlda"

Description
Idtlda contains the results of Linear Discriminant Analysis for the interval data

Slots
prior: Prior probabilities of class membership
means: Matrix with the mean vectors for each group
scaling: Matrix which transforms observations to discriminant functions, normalized so that within groups covariance matrix is spherical.
N: Number of observations

Methods
predict signature(object = "Idtlda"): Classifies interval-valued observations in conjunction with lda.
show signature(object = "Idtlda"): show S4 method for the IDdtlda-class

Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References

See Also
qda, MANOVA, IData
Description

IdtMANOVA contains the results of MANOVA tests on the interval-valued data.

Slots

nivar: Number of interval variables
grouping: Factor indicating the group to which each observation belongs to
H0res: Model estimates under the null hypothesis
H1res: Model estimates under the alternative hypothesis
QuiSq: Value of the Qui-Square statistics corresponding to the performed test
df: Degrees of freedom of the Qui-Square statistics
pvalue: p-value of the Qui-Square statistics value, obtained from the Qui-Square distribution with
        df degrees of freedom
H0logLik: Logarithm of the Likelihood function under the null hypothesis
H1logLik: Logarithm of the Likelihood function under the alternative hypothesis

Extends

Class "LRTest", directly.

Methods

show signature(object = "IdtMANOVA"): show $S4$ method for the IdtMANOVA-class

Author(s)

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References


See Also

MANOVA, lda, IData, IdtClMANOVA, IdtHetNMANOVA
Description

IdtMxE contains the results of a mixture model estimation.

Slots

grouping: Factor indicating the group to which each observation belongs to

ModelNames: The model acronym, indicating the model type, currently N for Normal followed by the configuration (C1 to C5)

ModelType: Indicates the model; currently, only Gaussian (Normal) distributions are implemented

ModelConfig: Configuration of the variance-covariance matrix: a vector of integers in the range 1 to 5

NIVar: Number of interval variables

SelCrit: The model selection criterion; currently, AIC and BIC are implemented

logLik: The logarithms of the likelihood function for the different cases

AICs: Value of the AIC criterion

BICs: Value of the BIC criterion

BestModel: The index of best model according to the chosen selection criterion

Extends

Class "IdtE", directly.

Author(s)

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References


See Also

MANOVA, IData
IdtMxNDE-class

Class "IdtMxNDE"

Description

IdtMxNDE contains the results of a mixture model estimation for the Normal model, with the five different possible variance-covariance configurations.

Slots

hmcdt: Indicates whether we consider an homocedastic (TRUE) or a hetereocedasic model (FALSE)
mlenmue: Matrix with the maximum likelihood mean vectors estimates by group (each row refers to a group)
mlenmuese: Matrix with the maximum likelihood means’ standard errors by group (each row refers to a group)
Configurations: List of the considered configurations
grouping: Factor indicating the group to which each observation belongs to
ModelNames: The model acronym, indicating the model type, currently N for Normal and the configuration (C1 to C5)
ModelType: Indicates the model; currently, only Gaussian (Normal) distributions are implemented
ModelConfig: Configuration of the variance-covariance matrix: a vector of integers in the range 1 to 5
NIVar: Number of interval variables
SelCrit: The model selection criterion; currently, AIC and BIC are implemented
logliks: The logarithms of the likelihood function for the different cases
AICs: Value of the AIC criterion
BICs: Value of the BIC criterion
BestModel: The index of best model according to the chosen selection criterion

Extends

Class "IdtMxE", directly. Class "IdtNDE", directly. Class "IdtE", by class "IdtMxE", distance 2.

Author(s)

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References


See Also

MANOVA, IData
IdtNDE-class

Class "IdtNDE"

Description

IdtNDE is a class union used to store the estimation results for Gaussian (Normal) models of Interval Data.

Methods

- **coef** signature(coef = "IdtNDE"): extracts parameter estimates from objects of type IdtNDE.
- **stdErr** signature(x = "IdtNDE"): extracts estimator standard errors from objects of type IdtNDE.

References


See Also

- `idata`, `mle`, `MANOVA`, `idtSngNDE`, `IdtMxNDE`

Idtqda-class

Class "Idtqda"

Description

Idtqda contains the results of Quadratic Discriminant Analysis for the interval data

Slots

- **prior**: Prior probabilities of class membership
- **means**: Matrix with the mean vectors for each group
- **scaling**: For each group i, scaling[,i] is an array which transforms interval-valued observations so that within-groups covariance matrix is spherical.
- **ldet**: Vector of half log determinants of the dispersion matrix.
- **lev**: Levels of the grouping factor

Methods

- **predict** signature(object = "Idtqda"): Classifies interval-valued observations in conjunction with qda.
- **show** signature(object = "Idtqda"): show S4 method for the Idtqda-class
IdtSngDE-class

Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

References

See Also

MANOVA, lda, IData

IdtSngDE-class  Class "IdtSngDE"

Description
IdtSngDE contains the results of a single class model estimation, with the five different possible variance-covariance configurations.

Slots

ModelNames: The model acronym, indicating the model type, currently N for Normal and the configuration (C1 to C5)
ModelType: Indicates the model; currently, only Gaussian (Normal) distributions are implemented
ModelConfig: Configuration of the variance-covariance matrix: a vector of integers in the range 1 to 5
NIVar: Number of interval variables
SelCrit: The model selection criterion; currently, AIC and BIC are implemented
logLiks: The logarithms of the likelihood function for the different cases
AICs: Value of the AIC criterion
BICs: Value of the BIC criterion
BestModel: The index of best model according to the chosen selection criterion

Extends

Class "IdtE", directly.

Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>
References

See Also
mle, IData

IdtSngNDE-class  

Class "IdtSngNDE"

Description
IdtSngNDE contains the results of a single class model estimation for the Normal distribution, with the five different possible variance-covariance configurations.

Slots
mleNmuE: Matrix with the maximum likelihood mean vectors estimates
mleNmuEse: Matrix with the maximum likelihood means’ standard errors by group
Configurations: List of the considered configurations
ModelNames: The model acronym, indicating the model type, currently N for Normal and the configuration (C1 to C5)
ModelType: Indicates the model; currently, only Gaussian (Normal) distributions are implemented
ModelConfig: Configuration of the variance-covariance matrix: a vector of integers in the range 1 to 5
NIVar: Number of interval variables
SelCrit: The model selection criterion; currently, AIC and BIC are implemented
logLik: The logarithms of the likelihood function for the different cases
AICs: Value of the AIC criterion
BICs: Value of the BIC criterion
BestModel: The index of best model according to the chosen selection criterion

Extends
Class "IdtSngDE", directly. Class "IdtNDE", directly. Class "IdtE", by class "IdtSngDE", distance 2.

Author(s)
Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>
References


See Also

mle, IData

Description

Linear Discriminant Analysis

Usage

```r
# S4 method for signature 'IdtClMANOVA'
lda(x, prior="proportions", selmodel=BestModel(H1res(x)), egvtol=1.0e-9,...)

# S4 method for signature 'IData'
lda(x, grouping, prior="proportions", tol=1.0e-4,
subset=1:nrow(x),Config=1:5, SelCrit=c("AIC","BIC"))
```

Arguments

- `x` An IdtClMANOVA object representing the results of a (homocedastic normal-based) MANOVA of Interval Data, or an IData object representing NObs interval-valued entities
- `grouping` Factor indicating the group to which each observation belongs to
- `prior` The prior probabilities of class membership. If unspecified, the class proportions for the training set are used. If present, the probabilities should be specified in the order of the factor levels
- `selmodel` The index of model of IdtClMANOVA object to be used in the construction of the discriminant functions
- `tol` A tolerance to decide if a matrix is singular; it will reject variables and linear combinations of unit-variance variables whose standard deviation is less than tol
- `egvtol` A tolerance for positive eigenvalues of the scatter matrix; eigenvalues below egvtol will be considered equal to 0, and the corresponding linear discriminant functions will be ignored
- `subset` An index vector specifying the cases to be used in the training sample. (NOTE: If given, this argument must be named
- `Config` Configuration of the variance-covariance matrix: a vector of integers between 1 and 5
- `SelCrit` The model selection criterion
- `...` Further arguments passed to or from other methods.
Value
An object of type Idtlda, containing the linear discriminant functions

Methods
signature(Idt = "IData") Performs Linear Discriminant Analysis on interval-valued data

See Also
Idtlda, DACrossVal, qda

Examples

# Create an Interval-Data object containing the intervals for 899 observations # on the temperatures by quarter in 60 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8])

# Linear Discriminant Analysis, assuming that Mid-Points are independent of LogRanges # (Configuration C4)
ChinaLdaC4 <- lda(ChinaT,ChinaTemp$GeoReg,Config=4)
cat("China linear discriminant analysis results =\n") ; print(ChinaLdaC4)
cat("Lda Prediction results: \n")
print(predict(ChinaLdaC4,ChinaT)$class)

# Estimate error rates by three-fold cross-validation, replicated five times
CVldaC4 <- DACrossVal(ChinaT,ChinaTemp$GeoReg,TrainAlg=lda,Config=4,kfold=3,CVrep=5)
summary(CVldaC4[,,"Clerr"])
glerrors <- apply(CVldaC4[,,"Nk"]*CVldaC4[,,"Clerr"],1,sum)/apply(CVldaC4[,,"Nk"],1,sum)
cat("Average global classification error =",mean(glerrors),"\n")

Description
LRTest contains the results of likelihood ratio tests

Slots
QuISq: Value of the Qui-Square statistics corresponding to the performed test
df: Degrees of freedom of the Qui-Square statistics
pvalue: p-value of the Qui-Square statistics value, obtained from the Qui-Square distribution with
df degrees of freedom
H0logLik: Logarithm of the Likelihood function under the null hypothesis
H1logLik: Logarithm of the Likelihood function under the alternative hypothesis
Methods

show signature(object = "LRTest"): show S4 method for the LRTest-class

Author(s)

Pedro Duarte Silva <psilva@porto.ucp.pt>
Paula Brito <mpbrito.fep.up.pt>

See Also

mle, IData, ConfTests, MANOVA

MANOVA-methods

Methods for Function MANOVA in Package ‘MAINT.Data’

Description

Performs MANOVA tests based on likelihood ratios allowing for both homocedastic and hetero-
cedastic setups

Usage

MANOVA(Idt, grouping, Model="Normal", Config=1:5,
SelCrit=c("AIC","BIC"), Mxt=c("Hom","Het"), tol=1.0e-4)

Arguments

Idt An IData object representing NObs interval-valued entities

grouping Factor indicating the group to which each observation belongs to

Model The joint distribution assumed for the MidPoint and LogRanges; currently, only
Gaussian (Normal) distributions are implemented

Config Configuration of the variance-covariance matrix: a vector of integers between 1 and 5

SelCrit The model selection criterion

Mxt Indicates whether we consider an homocedastic ("Hom") or a hetereocedasic ("Het") model configuration

tol Tolerance level for the standard deviation of non-constant variables. When a
MidPoint or LogRange has a within-groups standard deviation below tol, it is
considered to be a constant

Value

An object of type IdtMANOVA, containing the estimation and test results
Methods

signature(Idt = "IData") Performs MANOVA tests on interval-valued data

See Also

IdtMANOVA

Examples

# Create an Interval-Data object containing the intervals for 899 observations # on the temperatures by quarter in 60 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8], VarNames=c("T1","T2","T3","T4"))

# MANOVA tests, assuming that one of C3, C4 or C5 restricted configurations holds

# Classical (homocedastic) MANOVA tests
ManvChina <- MANOVA(ChinaT, ChinaTemp$GeoReg, Config=3:5)
cat("China, MANOVA by geografical regions results =\n")
print(ManvChina)

# Heterocedastic MANOVA tests
HetManvChina <- MANOVA(ChinaT, ChinaTemp$GeoReg, Mxt="Het", Config=3:5)
cat("China, heterocedastic MANOVA by geografical regions results =\n")
print(HetManvChina)

### mle-methods

Methods for Function mle in Package 'MAINT.Data'

Description

Performs maximum likelihood estimation for different distributions

Usage

mle(Idt, Model="Normal", Config=1:5, SelCrit=c("AIC","BIC"))

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idt</td>
<td>An IData object representing NObs interval-valued entities.</td>
</tr>
<tr>
<td>Model</td>
<td>The joint distribution assumed for the MidPoint and LogRanges; currently, only Gaussian (Normal) distributions are implemented</td>
</tr>
<tr>
<td>Config</td>
<td>Configuration of the variance-covariance matrix: a vector of integers between 1 and 5</td>
</tr>
<tr>
<td>SelCrit</td>
<td>The model selection criterion</td>
</tr>
</tbody>
</table>
Value

An object of type IdtSngDE, containing the estimation results

Methods

signature(Idt = "IData") Performs maximum likelihood estimation for interval-valued data, considering different distributions, and alternative variance-covariance matrix configurations

See Also

IdtSngDE

Examples

# Create an Interval-Data object containing the intervals for 899 observations
# on the temperatures by quarter in 60 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8])

# Estimate parameters by maximum likelihood, assuming that
# one of C3, C4 or C5 restricted configurations holds
ChinaE <- mle(ChinaT,Config=3:5)
cat("China maximum likelihood estimation results =\n") ; print(ChinaE)
cat("Standard Errors of Estimators:\n") ; print(stde(ChinaE))
prior The prior probabilities of class membership. If unspecified, the class proportions for the training set are used. If present, the probabilities should be specified in the order of the factor levels.

selmodel The index of model of IdtCIMANOVA object to be used in the construction of the discriminant functions.

tol A tolerance to decide if a matrix is singular; it will reject variables and linear combinations of unit-variance variables whose standard deviation is less than tol.

subset An index vector specifying the cases to be used in the training sample. (NOTE: If given, this argument must be named.

Config Configuration of the variance-covariance matrix: a vector of integers between 1 and 5.

SelCrit The model selection criterion.

... Further arguments passed to or from other methods.

Value
An object of type Idtqda, containing the quadratic discriminant functions.

Methods

signature(Idt = "IData") Performs Quadratic Discriminant Analysis on interval-valued data.

See Also
Idtllda, DACrossVal, lda.

Examples

# Create an Interval-Data object containing the intervals of temperatures
# by quarter for 60 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8])

# Create an Interval-Data object containing the intervals for 899 observations
# on the temperatures by quarter in 60 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8])

# Quadratic Discriminant Analysis, assuming independent Interval Variables
# (Configuration C3)
ChinaqdaC3 <- qda(ChinaT, ChinaTemp$GeoReg, Config=3)
cat("China quadratic discriminant analysis results =\n")
print(ChinaqdaC3)
cat("qda Prediction results: \n")
print(predict(ChinaqdaC3, ChinaT)$class)

# Estimate error rates by three-fold cross-validation, replicated five times
CVqdaC3 <- DACrossVal(ChinaT, ChinaTemp$GeoReg, TrainAlg=qda, Config=3, kfold=3, CVrep=5)
summary(CVqdaC3[,,"Clerr"])
glerrors <-
apply(CVqdaC3[,,"Nk"]*CVqdaC3[,,"Clerr"],1,sum)/apply(CVqdaC3[,,"Nk"],1,sum)
Methods for Function testMod in Package 'MAINT.Data'

Description

Performs statistical likelihood-ratio tests that evaluate the goodness-of-fit of a nested model against a more general one.

Usage

testMod(Idx, RestMod = 1:length(Idx@ModelNames), FullMod = "Next")

Arguments

- **Idx**: An IData object representing NObs interval-valued entities
- **RestMod**: Indices of the restricted models being evaluated in the NULL hypothesis
- **FullMod**: Either indices of the general models being evaluated in the alternative hypothesis or the strings "Next" (default) or "All". In the former case a Restricted model is always compared against the most parsimonious alternative that encompasses it, and in latter all possible comparisons are performed

Value

An object of type ConfTests with the results of the tests performed

See Also

ConfTests

Examples

```r
# Create an Interval-Data object containing the intervals of temperatures by quarter # for 899 Chinese meteorological stations.
ChinaT <- IData(ChinaTemp[1:8])

# Estimate by maximum likelihood the parameters of Gaussian models # with configurations C1, C3, C4 and C5 for the Winter (1st and 4th) quarter intervals #ChinaMTE <- mle(ChinaT[, c(1, 4)], Config = c(1, 3, 4, 5)) ChinaMTE <- mle(ChinaT[, c(1, 4)])
cat("China maximum likelihood estimation results for Winter quarters:
") print(ChinaMTE)

# Perform Likelihood-Ratio tests comparing models with consecutive nested Configuration
```
testMod(ChinaWTE)

# Perform Likelihood-Ratio tests comparing all possible models
testMod(ChinaWTE,FullMod="All")

# Compare model with covariance Configuration 4
# against model with covariance Configuration 1 (unrestricted covariance)
testMod(ChinaWTE,RestMod=4,FullMod=1)
Index

+Topic **AIC**  
  BestModel-methods, 4

+Topic **BIC**  
  BestModel-methods, 4

+Topic **Interval Data**  
  IdtNDE-class, 17  
  MAINT.Data-package, 2

+Topic **Linear Discriminant Analysis**  
  lda-methods, 20

+Topic **MAINT.Data**  
  MAINT.Data-package, 2

+Topic **MANOVA for Interval Data**  
  MAINT.Data-package, 2

+Topic **MANOVA**  
  IdtC1MANOVA-class, 10  
  IdtHetNMANOVA-class, 12  
  IdtMANOVA-class, 14  
  MANOVA-methods, 22

+Topic **Parametric modelling of Interval Data**  
  MAINT.Data-package, 2

+Topic **Quadratic Discriminant Analysis**  
  qda-methods, 24

+Topic **Statistical tests for Interval Data**  
  MAINT.Data-package, 2

+Topic **Symbolic Data Analysis**  
  MAINT.Data-package, 2

+Topic **classes**  
  ConfTests-class, 5  
  IData-class, 9  
  IdtC1MANOVA-class, 10  
  IdtE-class, 11  
  IdtHetNMANOVA-class, 12  
  Idtlda-class, 13  
  IdtMANOVA-class, 14  
  IdtMxE-class, 15  
  IdtMXNDE-class, 16  
  IdtNDE-class, 17  
  Idtqda-class, 17  
  IdtSngDE-class, 18  
  IdtSngNDE-class, 19  
  LRTest-class, 21

+Topic **cross-validation**  
  DACrossVal, 6

+Topic **datasets**  
  ChinaTemp, 5

+Topic **discriminant analysis**  
  DACrossVal, 6

+Topic **interval data**  
  BestModel-methods, 4  
  ChinaTemp, 5  
  ConfTests-class, 5  
  IData, 8  
  IData-class, 9  
  IdtC1MANOVA-class, 10  
  IdtE-class, 11  
  IdtHetNMANOVA-class, 12  
  Idtlda-class, 13  
  IdtMANOVA-class, 14  
  IdtMxE-class, 15  
  IdtMXNDE-class, 16  
  Idtqda-class, 17  
  IdtSngDE-class, 18  
  IdtSngNDE-class, 19  
  lda-methods, 20  
  LRTest-class, 21  
  MANOVA-methods, 22  
  mle-methods, 23  
  qda-methods, 24  
  testMod-methods, 26

+Topic **likelihood ratio test**  
  LRTest-class, 21  
  testMod-methods, 26

+Topic **likelihood ratio tests**  
  MANOVA-methods, 22

+Topic **maximum likelihood**
estimation
mle-methods, 23
*Topic methods
  BestModel-methods, 4
  lda-methods, 20
  MANOVA-methods, 22
  mle-methods, 23
  qda-methods, 24
  testMod-methods, 26

BestModel (BestModel-methods), 4
BestModel, IdtE-method (IdtE-class), 11
BestModel-methods, 4

ChinaTemp, 5
ConfTests, 22, 26
ConfTests-class, 5

DACrossVal, 6, 21, 25
	head, IData-method (IData-class), 9

IData, 6, 7, 8, 10–20, 22
IData-class, 9
IdtClMANOVA, 14
IdtClMANOVA-class, 10
IdtE, 15, 16, 18, 19
IdtE-class, 11
IdtHetNMANOVA, 14
IdtHetNMANOVA-class, 12
Idtlda, 21, 25
Idtlda-class, 13
IdtMANOVA, 10, 12, 23
IdtMANOVA-class, 14
IdtMxE, 16
IdtMxE-class, 15
IdtMxNDE, 17
IdtMxNDE-class, 16
IdtNDE, 16, 19
IdtNDE-class, 17
Idtqda-class, 17
IdtSngDE, 19, 24
IdtSngDE-class, 18
IdtSngNDE, 17
IdtSngNDE-class, 19

lda, 7, 10, 11, 14, 18, 25
lda (lda-methods), 20
lda, IData-method (lda-methods), 20
lda, IdtClMANOVA-method (lda-methods), 20

lda-methods, 20
LRTest, 6, 10, 12, 14
LRTest-class, 21

MAINT.Data (MAINT.Data-package), 2
MAINT.Data-package, 2
MANOVA, 10–18, 22
MANOVA (MANOVA-methods), 22
MANOVA, IData-method (MANOVA-methods), 22
MANOVA-methods, 22
mle, 6, 10, 12, 17, 19, 20, 22
mle (mle-methods), 23
mle, IData-method (mle-methods), 23
mle-methods, 23

predict, Idtlda-method (Idtlda-class), 13
predict, Idtqda-method (Idtqda-class), 17
print, IData-method (IData-class), 9

qda, 7, 10, 13, 21
qda (qda-methods), 24
qda, IData-method (qda-methods), 24
qda, IdtHetNMANOVA-method (qda-methods), 24
qda-methods, 24

show, ConfTests-method
  (ConfTests-class), 5
show, IdtE-method (IdtE-class), 11
show, Idtlda-method (Idtlda-class), 13
show, IdtMANOVA-method
  (IdtMANOVA-class), 14
show, Idtqda-method (Idtqda-class), 17
show, LRTest-method (LRTest-class), 21
summary, IData-method (IData-class), 9

tail, IData-method (IData-class), 9
testMod (testMod-methods), 26
testMod, IdtE-method (IdtE-class), 11
testMod-methods, 26