Package 'incReg'

July 2, 2014

Type Package

Title Incremental Multivariate Regression

Version 1.0

Date 2013-05-04

Maintainer Alexander Entzian <a.entzian@alexander-entzian.de>

Description The incReg package calculate the incremental multivariate Regression based on a given dataset or a .csv file, based on the multiple linear regression

License LGPL (>= 3)

Depends R (>= 2.14.0), car, methods

LazyLoad yes

Author Alexander Entzian [cre, aut, cph]

NeedsCompilation no

Repository CRAN

Date/Publication 2013-10-10 09:50:24

R topics documented:

`incReg-package` *incremental multivariate Regression*

Description

The incReg package calculate the incremental multivariate Regression based on a given dataset or a .csv file, based on the multiple linear regression

Details

Package:	incReg
Type:	Package
Version:	1.0
Date:	2013-05-05
License:	LGPL (>= 3)
Depends:	R (>= 2.14.0), car, methods

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

`incReg` `IncReg-class`

Examples

```
data(exampleIncRegData)

result <- incReg(yName = "exp.pkb", dataSet = exampleIncRegData)
```

```
summary(result)
finalModel <- getFinalModel(result)
```

`checkAndSetDefaultValues` *check and set default values*

Description

checked some parameters if they are in their specified limits. Otherwise, the default value is used.

Usage

```
checkAndSetDefaultValues(listOfData)
```

Arguments

`listOfData`	should be a list of the parameter or a object of the class `call`

Details

The values to be checked are currently: csvFile, missingValueTyp, cutlineZeroTest, cutlineAdvZeroTest , cutlineCorr, cutlineMulticoll, alphaValue

Value

the returned value is a list of the adapted values

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

`incReg`

convertAllToNumeric *convert all to numeric values*

Description

convert all not numeric values to numeric or NA thereby character convert to NAs also empty values like ” will be convert to NAs

Usage

```
convertAllToNumeric(dataSet, ...)
```

Arguments

dataSet	any data set
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

correlationTest *correlation Test*

Description

Check if a correlation exits between two descriptors

Usage

```
correlationTest(dataSet, yName, cutlineCorr, ...)
```

Arguments

dataSet	Input data set (should be numeric values)
yName	Name of the column which should be predicted
cutlineCorr	The column, from a correlation pair (value more than the cutline), which has less correlation agains the y column is removed
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

car, incReg

---

deletRowsAndColumsAndOrCalculatMean
delet rows and colums and/or calculat mean

Description

Delete all rows and cols with NA or calculate the mean for each cell (special handling for the y column)

Usage

```
deletRowsAndColumsAndOrCalculatMean(dataSet, yName,
missingValueTyp, missingValueIsCol, ...)
```

Arguments

dataSet	input data set
yName	Name of the column which should be predicted
missingValueTyp	remove ("rm") or average ("avg")
missingValueIsCol	if missingValueTyp is "rm" then this value is important: TRUE -> column will be removed, FALSE -> the row
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

`exampleIncRegData` *experimental pkb values*

Description

example data set for the package incReg. contents 36 substances with experimental pkb values and 6 descriptors calculated with chemoinformatics methods (MSINDO, PM6, GAUSSIAN)

Usage

```
data(exampleIncRegData)
```

Format

A data set with 36 substances and 8 columns

`Name` a character vector

`exp pkb` a numeric vector

`dipole D` a numeric vector

`total energy eV` a numeric vector

`Klopman one divided by loca au` a numeric vector

`Loca energy eV` a numeric vector

`charge of N e` a numeric vector

`proton affinity kcal/mol` a numeric vector

References

Sanderson P (2003) Azaindoles: moderately basic P1 groups for enhancing the selectivity of thrombin inhibitors. Bioorg Med Chem Lett 13:795-798. doi: 10.1016/S0960-894X(03)00017-9 Perrin DD (1965) Prediction of the strengths of some organic bases. J Chem Soc 18:5590-5596. doi: 10.1039/jr9650005590 Albert A, Goldacre R, Phillips J (1948) 455. The Strength of Heterocyclic Bases. J Chem Soc 2240-2249. doi: 10.1039/JR9480002240 Catalan J, Elguero J (1983) Basicity of azoles. Part 2. Theoretical study of the basicity of methylpyrazoles and methylimidazoles. J Chem Soc Perkin Trans II 1869. doi: 10.1039/p29830001869 Catalan J, Menendez M, Laynez J, Claramunt RM, Bruix M, Mendoza J De, Elguero J (1985) Basicity of azoles. VII. Basicity of C-aminopyrazoles in relation to tautomeric and protonation studies. J Heterocyclic Chem 22:997-1000. doi: 10.1002/jhet.5570220413 Kostka K, Strawiak MM (1984) Protolytic and complexing properties of pyrazoles. Part III. 1-Phenylpyrazole, 3(5)-phenylpyrazole, 3,5-diphenylpyrazole, 1-methyl-3,5-dimethylpyrazole, and 1,3,5-triphenylpyrazole. Pol J Chem 58:647 Llamas-Saiz AL, Foces-Foces C, Cano FH, Jimenez P, Laynez J, Meutermans W, Elguero J, Limbach H-H, Aguilar-Parrilla F (1994) The influence of the nitro group on the solid-state structure of 4-nitropyrazoles: the cases of pyrazole, 3,5-dimethylpyrazole, 3,5-di- tert -butylpyrazole and 3,5-diphenylpyrazole. I. Static aspects (crystallography and thermodynamics). Acta Crystallogr B 50:746-762. doi: 10.1107/S0108768194004180 Jiang F, McCracken J, Peisach J (1990) Nuclear quadrupole interactions in copper(II)-diethylenetriamine-substituted imidazole complexes and in copper(II) proteins. J

Am Chem Soc 112:9035-9044. doi: 10.1021/ja00181a002 Takeuchi Y, Kirk KL, Cohen LA (1978) Adjacent lone pair (ALP) effects in heteroaromatic systems. 2. Isotope exchange of ring hydrogens in nitro- and fluoroimidazoles. J Org Chem 43:3570-3578. doi: 10.1021/jo00412a034 ACD/Labs (2009) ACD/pKa. http://www.acdlabs.com/products/phys_chem_lab/pka/exp_table.html. Accessed 02 May 2009 Cassidy CS, Reinhardt LA, Cleland WW, Frey PA (1999) Hydrogen bonding in complexes of carboxylic acids with 1-alkylimidazoles: steric and isotopic effects on low barrier hydrogen bonding. J Chem Soc Perkin Trans II 635-642. doi: 10.1039/a806387g

---

`getAlphaValue` *Getter functions*

Description

Returns the corresponding value from an incReg-Object

Usage

```
getAlphaValue(object)
```

Arguments

`object`	represents an object of the class incReg

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

`setAlphaValue<-` `incReg`

`getModelNumberX` *Get result models*

Description

Return the result models of the incremental Multivariate Regression

Usage

```
getModelNumberX(object, number)

getFinalModel(object)
```

Arguments

object	represents an object of the class incReg
number	number of the model which is produced during the incremental Multivariate Regression (last number is normally the final model)

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

---

getPreModelResults *Multicollinearity and correlation results*

Description

Return the results of the multicollinearity and the correlation test

Usage

```
getPreModelResults(object)
```

Arguments

object	represents an object of the class incReg

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

incReg *incReg constructor*

Description

Constructor for users. Creates the incReg-object and start the process.

Usage

```
incReg(yName, dataSet = data.frame(), csvFile = "None", doAnalysis = TRUE,
   shouldScaling = FALSE, shouldCentering = FALSE, missingValueTyp = "rm",
   missingValueIsCol = TRUE, cutlineZeroTest = 0.8, cutlineAdvZeroTest = 0.25,
   cutlineCorr = 0.49, cutlineMulticoll = 10, alphaValue = 0.05)
```

Arguments

yName	column name of the column which is to be predicted
dataSet	input data set, alternative see csvFile
csvFile	input csv File, alternative see dataSet
doAnalysis	should the process be performed directly (TRUE or FALSE)
shouldScaling	should the data set be scaled (TRUE or FALSE)
shouldCentering	should the data set be centered (TRUE or FALSE)
missingValueTyp	what should the method do with missing values: remove ("rm") or calculate the average ("avg")
missingValueIsCol	if missingValueTyp is "rm" then this value is important: TRUE -> column will be removed, FALSE -> the row
cutlineZeroTest	not more than the given cutline should be zero else the descriptor have to be removed
cutlineAdvZeroTest	there have to be no descriptor which have more than the given cutline on the same cases
cutlineCorr	the column, from a correlation pair (value more than the cutline), which has less correlation against the y column is removed
cutlineMulticoll	if a descriptor have a higher value than this cutline, multicollinearity is there
alphaValue	significance level

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg-package IncReg-class

Examples

```
data(exampleIncRegData)

result <- incReg(yName = "exp.pkb", dataSet = exampleIncRegData)
```

IncReg-class *Class* "IncReg"

Description

The incReg class calculate the incremental multivariate Regression based on a given dataset or a .csv file, based on the multiple linear regression

Objects from the Class

Objects can be created by calls of the form `new("IncReg", ...)`.

Slots

.dataSet: Object of class "data.frame" ~~

.csvFile: Object of class "character" ~~

.yName: Object of class "character" ~~

.shouldScaling: Object of class "logical" ~~

.shouldCentering: Object of class "logical" ~~

.missingValueTyp: Object of class "character" ~~

.missingValueIsCol: Object of class "logical" ~~

.cutlineZeroTest: Object of class "numeric" ~~

.cutlineAdvZeroTest: Object of class "numeric" ~~

.cutlineCorr: Object of class "numeric" ~~

.cutlineMulticoll: Object of class "numeric" ~~

.alphaValue: Object of class "numeric" ~~

.multiCollResult: Object of class "list" ~~

.collResult: Object of class "matrix" ~~

.modelResult: Object of class "list" ~~

Extends

Class "IncRegOption", directly. Class "IncRegFinalmodelResult", directly.

Methods

[signature(x = "IncReg"): ...

[<- signature(x = "IncReg"): ...

getAlphaValue signature(object = "IncReg"): ...

getCollResult signature(object = "IncReg"): ...

getCsvFile signature(object = "IncReg"): ...

getCutlineAdvZeroTest signature(object = "IncReg"): ...

getCutlineCorr signature(object = "IncReg"): ...

getCutlineMulticoll signature(object = "IncReg"): ...

getCutlineZeroTest signature(object = "IncReg"): ...

getDataSet signature(object = "IncReg"): ...

getFinalModel signature(object = "IncReg"): ...

getMissingValueIsCol signature(object = "IncReg"): ...

getMissingValueTyp signature(object = "IncReg"): ...

getModelNumberX signature(object = "IncReg"): ...

getModelResult signature(object = "IncReg"): ...

getMultiCollResult signature(object = "IncReg"): ...

getPreModelResults signature(object = "IncReg"): ...

getShouldCentering signature(object = "IncReg"): ...

getShouldScaling signature(object = "IncReg"): ...

getSquareBracket signature(x = "IncReg"): ...

getYname signature(object = "IncReg"): ...

print signature(x = "IncReg"): ...

processing signature(IncReg = "IncReg"): ...

setAlphaValue<- signature(object = "IncReg"): ...

setCollResult<- signature(object = "IncReg"): ...

setCsvFile<- signature(object = "IncReg"): ...

setCutlineAdvZeroTest<- signature(object = "IncReg"): ...

setCutlineCorr<- signature(object = "IncReg"): ...

setCutlineMulticoll<- signature(object = "IncReg"): ...

setCutlineZeroTest<- signature(object = "IncReg"): ...

setDataSet<- signature(object = "IncReg"): ...

setMissingValueIsCol<- signature(object = "IncReg"): ...

setMissingValueTyp<- signature(object = "IncReg"): ...

setModelResult<- signature(object = "IncReg"): ...

setMultiCollResult<- signature(object = "IncReg"): ...

setShouldCentering<- signature(object = "IncReg"): ...

setShouldScaling<- `signature(object = "IncReg")`: ...

setSquareBracket `signature(x = "IncReg")`: ...

setYname<- `signature(object = "IncReg")`: ...

show `signature(object = "IncReg")`: ...

summary `signature(object = "IncReg")`: ...

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg incReg-package

Examples

```
showClass("IncReg")
```

IncRegFinalmodelResult-class

Class "IncRegFinalmodelResult"

Description

Save the results of the incReg class

Objects from the Class

Objects can be created by calls of the form `new("IncRegFinalmodelResult", ...)`.

Slots

`.multiCollResult:` Object of class "list" ~~

`.collResult:` Object of class "matrix" ~~

`.modelResult:` Object of class "list" ~~

Methods

No methods defined with class "IncRegFinalmodelResult" in the signature.

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

Examples

```
showClass("IncRegFinalmodelResult")
```

IncRegOption-class *Class* "IncRegOption"

Description

save the options of the incReg class

Objects from the Class

Objects can be created by calls of the form `new("IncRegOption", ...)`.

Slots

`.shouldScaling`: Object of class `"logical"` ~~

`.shouldCentering`: Object of class `"logical"` ~~

`.missingValueTyp`: Object of class `"character"` ~~

`.missingValueIsCol`: Object of class `"logical"` ~~

`.cutlineZeroTest`: Object of class `"numeric"` ~~

`.cutlineAdvZeroTest`: Object of class `"numeric"` ~~

`.cutlineCorr`: Object of class `"numeric"` ~~

`.cutlineMulticoll`: Object of class `"numeric"` ~~

`.alphaValue`: Object of class `"numeric"` ~~

Methods

No methods defined with class "IncRegOption" in the signature.

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

Examples

```
showClass("IncRegOption")
```

incrementalMultivariateRegression *incremental Multivariate Regression*

Description

processed the incremental Multivariate Regression

Usage

```
incrementalMultivariateRegression(dataSet, yName, alphaValue, ...)
```

Arguments

dataSet	input data set
yName	Name of the column which should be predicted
alphaValue	significance level
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg car stats

multicollinearityTest *multicollinearity Test*

Description

Check if a multicollinearity exits between more than two descriptors. If more than two descriptors have multicol. then the descriptor with the lowest correlation to "y" removed. After that multicollinearity is recalculated.

Usage

```
multicollinearityTest(dataSet, yName, cutlineMulticoll, ...)
```

Arguments

dataSet	input data set
yName	Name of the column which should be predicted
cutlineMulticoll	if a descriptor have a higher value than this cutline, multicollinearity is there
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

`processing` *Miscellaneous*

Description

miscellaneous functions no use for external

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

`reanalyse` *Reanalyse*

Description

starts the process afresh

Usage

```
reanalyse(object)
```

Arguments

`object`	represents an object of the class incReg

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

scalingAndCenteringData *scaling and centering*

Description

Scaling and centering of the dataset

Usage

```
scalingAndCenteringData(dataSet, shouldScaling, shouldCentering, ...)
```

Arguments

dataSet	input data set
shouldScaling	if the data should scaled: TRUE or FALSE
shouldCentering	if the data should centered: TRUE or FALSE
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

setAlphaValue<- *Setter functions*

Description

Set the corresponding value for an incReg-Object

Usage

```
setAlphaValue(object) <- value
```

Arguments

object	represents an object of the class incReg
value	represents the corresponding value which should change or set

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

getAlphaValue incReg

---

summary *Summary*

Description

summary of the model(s)

Usage

```
summary(object)
```

Arguments

object	represents an object of the class incReg

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

---

zeroAndAdvancedZeroTest
zero and advanced zero test

Description

Check if the single columns have enough variations

Usage

```
zeroAndAdvancedZeroTest(dataSet, yName, cutlineZeroTest, cutlineAdvZeroTest, ...)
```

Arguments

dataSet	input data set
yName	Name of the column which should be predicted
cutlineZeroTest	not more than the given cutline should be zero
cutlineAdvZeroTest	there have to be no descriptor which have more than the given cutline on the same cases
...	not used

Author(s)

Alexander Entzian <a.entzian@alexander-entzian.de>

See Also

incReg

Index