

Package ‘AutoTransQF’

September 15, 2023

Type Package

Title A Novel Automatic Shifted Log Transformation

Version 0.1.3

Date 2023-09-13

Description A novel parametrization of log transformation and a shift parameter to automate the transformation process are proposed in R package 'AutoTransQF' based on Feng et al. (2016). Please read Feng et al. (2016) <[doi:10.1002/sta4.104](https://doi.org/10.1002/sta4.104)> for more details of the method.

License GPL-3

URL <https://github.com/yyuehhu/AutoTransQF>

Imports VGAM, matlab2r, moments, stats (>= 4.2.2)

Suggests nortest

LazyData true

RoxygenNote 7.2.3

Encoding UTF-8

Depends R (>= 3.5)

NeedsCompilation no

Author Yue Hu [aut, cre],
Hyeon Lee [aut],
J. S. Marron [aut]

Maintainer Yue Hu <yuehu2000@outlook.com>

Repository CRAN

Date/Publication 2023-09-15 19:14:40 UTC

R topics documented:

AutoTransQF-package	2
ADStatQF	2
autotransfuncQF	3
AutoTransQF	4
Melanoma	5

Index[7](#)

AutoTransQF-package *A Novel Automatic Shifted Log Transformation*

Description

The R package AutoTransQF based on Feng et al.(2016) introduces a novel parametrization of log transformation and a shift parameter to automate the transformation process. Please read Feng et al. (2016) <doi: 10.1002/sta4.104> for more details of the method.

Author(s)

Yue Hu [aut, cre], Hyeon Lee [aut], J. S. Marron [aut]

ADStatQF *Calculates Anderson-Darling Test Statistic*

Description

The function is used to calculate the Anderson-Darling test statistic of standard normal distribution. The input needs to be vectors with seven or more entries.

Usage

```
## Calculate Anderson-Darling test statistic for vector x:  
ADStatQF(x)
```

Arguments

x a vector with seven or more entries.

Value

Returns the Anderson-Darling test statistic for the given vector.

Author(s)

Yue Hu, Hyeon Lee, J. S. Marron

References

Feng, Q. , Hannig J. , Marron, J. S. (2016). A Note on Automatic Data Transformation. STAT, 5, 82-87. doi: 10.1002/sta4.104

See Also

[autotransfuncQF](#), [AutoTransQF](#)

Examples

```
## Generate a vector z from normal distribution with mean 0 and standard deviation 6.
z = rnorm(10, mean = 0, sd = 6)

## Calculate the Anderson-Darling test statistic for z
ADStatQF(z)
```

autotransfuncQF *Transforms Individual Vectors into Normality*

Description

This function transforms individual vectors into normality. Paper from Feng et al. (2016) includes more details about the transformation mechanism.

Usage

```
autotransfuncQF(vari, istat, paraindex)
```

Arguments

vari	a vector needs to be transformed
istat	a value representing the type of test statistic for evaluation of normality of the transformed vector. If istat = 1, Anderson-Darling test statistic is chosen; if istat = 2, standard skewness statistic is chosen.
paraindex	a value delivered to calculate the specific shift parameter beta

Value

Returns a transformed vector with the shift parameter calculated by the input paraindex

Author(s)

Yue Hu, Hyeon Lee, J. S. Marron

References

Feng, Q. , Hannig J. , Marron, J. S. (2016). A Note on Automatic Data Transformation. STAT, 5, 82-87. doi: 10.1002/sta4.104

See Also

[ADStatQF](#), [AutoTransQF](#)

Examples

```
## Generate a vector vec from Gamma distribution with parameters shape 1 and scale 2.
vec = rgamma(50, shape = 1, scale = 2)

## Choose Anderson-Darling test statistic for transformed vector.
## Assign paraindex to be 0.9.
autotransfuncQF(vec, istat = 1, paraindex = 0.9)

## Choose skewness for transformed vector
## Assign paraindex to be 0.9
autotransfuncQF(vec, istat = 2, paraindex = 0.9)
```

AutoTransQF

Transforms Datasets into Normality

Description

This function helps to transform each vector of the matrix into normality based on the optimal test statistic of transformed vectors.

Usage

```
## The function tries to transform each vector of mdata into normality
AutoTransQF(mdata, paramstruct = list(istat, iscreenwrite, FeatureNames))
```

Arguments

mdata	the matrix needs to be transformed.
paramstruct	A list with three entries istat, iscreenwrite and FeatureNames respectively. Missing entries will be set to default.
istat	a value representing the type of test statistic for evaluation of normality of the transformed vector with default to be istat = 1. If istat = 1, Anderson-Darling test statistic is chosen; if istat = 2, standard skewness statistic is chosen.
iscreenwrite	Whether there is screenwrite with default to be iscreenwrite = 0. If iscreenwrite = 1, to write progress to screen; if iscreenwrite = 0, no screenwrite.
FeatureNames	Contains feature names of each vector with default to be 'Feature1'

Value

Returns a list with three elements:

data	the transformed matrix
beta	a list of all shift parameters beta
alpha	a list of all shift parameters alpha

Note

When a vector of the original matrix is not transformed, its corresponding alpha and beta are both -1.

Author(s)

Yue Hu, Hyeon Lee, J. S. Marron

References

Feng, Q. , Hannig J. , Marron, J. S. (2016). A Note on Automatic Data Transformation. *STAT*, 5, 82-87. doi: 10.1002/sta4.104

See Also

[ADStatQF](#), [autotransfuncQF](#)

Examples

```
## Create a random matrix x.
x = matrix(rgamma(40, shape = 1, scale = 2), nrow = 4)

## Transform matrix x in default setting and
## output transformed data
AutoTransQF(x)$data

## Transform matrix x in default setting and
## output a list of shift parameter beta
AutoTransQF(x)$beta

## Transform matrix x with feature names and
## output a list of shift parameter alpha
Names = c('Feature1', 'Feature2', 'Feature3', 'Feature4')
AutoTransQF(x, paramstruct = list(FeatureNames = Names))$alpha

## Transform matrix x with feature names, progress to screen,
## and apply standard skewness statistic to transformed vectors
AutoTransQF(x, paramstruct = list(istat = 2, iscreenwrite = 1, FeatureNames = Names))

## Transform matrix x with progress to screen and
## apply standard skewness statistic to transformed vectors
AutoTransQF(x, paramstruct = list(istat = 2, iscreenwrite = 1))
```

Melanoma

MelanomaFeatures from Miedema et al. (2012)

Description

The dataset from Miedema et al. (2012) is built based on 49 hematoxylin and eosin stained slides of distinctive melanocytic lesions.

Usage

Melanoma

Format

A data frame with 10152 observations on 49 variables where columns serve as data objects and rows serve as features. All 49 features are numeric variables including Area, Hu . 1, Hu . 2 etc.

Source

Miedema, Jayson, et al. (2012). Image and statistical analysis of melanocytic histology. *Histopathology*, 61(3), pp.436-444. doi: 10.1111/j.1365-2559.2012.04229.x

Index

* datasets

Melanoma, [5](#)

ADStatQF, [2](#), [3](#), [5](#)

autotransfuncQF, [2](#), [3](#), [5](#)

AutoTransQF, [2](#), [3](#), [4](#)

AutoTransQF-package, [2](#)

Melanoma, [5](#)