Package 'codecountR'

December 7, 2023

Title Counting Codes in a Text and Preparing Data for Analysis

Version 0.0.4.0

Description Data analysis frequently requires coding, in particular when data are collected by interviews, by observations or even by questionnaires. Therefore, code counting and data preparation are necessary phases to carry out the analysis. Thus, the analysts will wish to count the codes inserted in a text (tokenization and counting of a list of pre-established codes) and to carry out the preparation of the data (feature scaling min-max normalization, Zscore, Box and Cox transformation, non parametric bootstrap). For Box and Cox (1964) <https://www.jstor.org/stable/2984418> transformation, optimal Lambda is calculated by log-likelihood. Non parametric bootstrap is based on randomly sampling data with replacement. Package for educational purposes.

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

Imports stats

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Philippe Cohard [aut, cre]

Maintainer Philippe Cohard <p.cohard@laposte.net>

Repository CRAN

Date/Publication 2023-12-07 15:10:02 UTC

R topics documented:

---

analysCodesList   *analysCodesList*

Description

analysCodesList

Usage

```
analysCodesList(dataS, codesLis)
```

Arguments

dataS	a character
codesLis	a character

Value

a list

Examples

```
codes=list("@essai@","@test@")
data = "this is an example @essai@, a bit long @essai@ text"
Result=analysCodesList(data,codes)
Result
```

---

bootStrap   *bootStrap*

Description

bootStrap

Usage

```
bootStrap(nameDframe, grpSize)
```

Arguments

nameDframe	a data.frame
grpSize	a number

Value

a matrix

Examples

```
j=c(10,14,56,30,58,78,99,1)
k=c(10,12,14,16,18,20,22,24)
x=data.frame(j,k)
res=bootStrap(x,5)
res
```

BoxAndCox   *BoxAndCox*

Description

BoxAndCox

Usage

```
BoxAndCox(rawVect, minLambda)
```

Arguments

rawVect   a vector

minLambda   a number

Value

a list

Examples

```
vec=rlnorm(100, log(3), log(3))
BandC=BoxAndCox(vec, -3)
BandC
BAC=unlist(BandC$par)
BAC
rawVectBCFinal=unlist(subCalcBoxAndCox(vec, BandC$par))
```

---

codeCount *codeCount*

Description

codeCount

Usage

```
codeCount(dataSet, code)
```

Arguments

dataSet	a character
code	a character

Value

a number

Examples

```
data = "this is an example @essai@"
codeCount(data, "@essai@") #number of lines containing the chain
```

---

loadCodes *loadCodes*

Description

loadCodes

Usage

```
loadCodes(txtFile)
```

Arguments

txtFile	a character

Value

a list

Examples

```
theFile =system.file("codesList.txt", package = "codecountR")
data=loadCodes(theFile)
```

normMinMax *normMinMax*

Description

normMinMax

Usage

```
normMinMax(nameDframe)
```

Arguments

nameDframe a data.frame

Value

a data.frame

Examples

```
j=c(10,14,56,30,58,78,99,1)
k=c(10,12,14,16,18,20,22,24)
x=data.frame(j,k)
xMinMax=normMinMax(x)
xMinMax
```

subCalcBoxAndCox *subCalcBoxAndCox*

Description

subCalcBoxAndCox

Usage

```
subCalcBoxAndCox(sortedVect, actualLambda)
```

Arguments

sortedVect a vector

actualLambda a number

Value

a vector

Examples

```
vec=rlnorm(100, log(3), log(3))
BandC=subCalcBoxAndCox(vec, -3)
```

---

tokenization   *tokenization*

Description

tokenization

Usage

```
tokenization(txtFile)
```

Arguments

txtFile   a character

Value

a list

Examples

```
theFile =system.file("ExText.txt", package = "codecountR")
data=tokenization(theFile)
```

---

zScore   *zScore*

Description

zScore

Usage

```
zScore(nameDframe)
```

Arguments

nameDframe   a data.frame

Value

a data.frame

Examples

```
j=c(10,14,56,30,58,78,99,1)
k=c(10,12,14,16,18,20,22,24)
x=data.frame(j,k)
xZsc=zScore(x)
xZsc
```

Index