Package 'RDA'

February 25, 2017

Type Package

Title Datasets, functions and examples from the book: R Data Analysis-Methods and Application (in chinese)by Kuangnan Fang et al

Version 0.1

Date 2017-02-19

Author Zhao Xue

Maintainer Zhao Xue <15420161152187@stu.xmu.edu.cn>

Description The collection of datasets, functions used in the book ``R Data Analysis-Methods and Application''.

License GPL(>2)

LazyData TRUE

RoxygenNote 5.0.1

Depends R (>= 2.10)

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

---

click   *Dataset of Web hits in chapter15*

Description

A dataset containing the length of keywords,display times,price of hits and average grade of 40 round objects.

Usage

```
click
```

Format

A data frame with 44 rows and 5 variables:

click the quantity of web hits

key_length the length of the keywords

display display times

price average price of web hits

grade the average rank

---

consumption_income   *Dataset of income and consumption in chapter12*

Description

A dataset containing year,consumption and income 2 variables of 36 objects

Usage

```
consumption_income
```

Format

a dataframe with 36 rows and 3 variables

Y year

C consumption

Y income

CTG   *Dataset of fetal state*

Description

A dataset containing LB,AC,FM...23 variables of 2126 obejects

Usage

```
CTG
```

Format

a dataframe with 2126 rows and 23 variables

LB FHR basic line:heart rate per minute

AC Accumulate times per second

FM Fetal movement per second

UC uterine contraction per second

DL Light deceleration per second

DS Serious deceleration per second

DP Persistent deceleration per second

ASTV Short term variation in the percentage of time

MSTV Short term variation of average

ALTV Long term variation in the percentage of time

MLTCV Long term variation of average

Width Width of FHR histogram

Min The minimum value of FHR histogram

Max The maximum value of FHR histogram

Nmax The peak count of FHR histogram

Nzeros The zero count of FHR histogram

Mode The mode of FHR histogram

Mean The mean of FHR histogram

Median The median of FHR histogram

Variance The variance of FHR histogram

Tendency The tendency of FHR histogram

CLASS The class of FHR

NSP The class of fetal state

Source

http://archive.ics.uci.edu/ml/datasets/Cardiotocography

---

douglas	*Dataset of argiculture production in chapter12*

Description

A dataset containing y,x2,x3 of 15 objects

Usage

```
douglas
```

Format

a dataframe with 15 rows and 3 variables

y actual output

x2 labor time

x3 actual capital input

EDA *Function of 'Exploratory Data Analysis'*

Description

Function of 'Exploratory Data Analysis'

Usage

```
EDA(x)
```

Arguments

x   Numeric vectors

Value

The outputs will be several graphs to describe some features of our data.

Examples

```
pay <- c(11,19,14,22,14,28,13,81,12,43,11,16,
31,16,23,42,22,26,17,22,13,27,108,16,43,82,
14,11,51,76,28,66,29,14,14,65,37,16,37,35,
39,27,14,17,13,38,28,40,85,32,25,26,16,12,
54,40,18,27,16,14,33,29,77,50,19,34)
EDA(pay)
log.pay <- log10(pay)
EDA(log.pay)
```

---

fair *Dataset of fairmodel study in chapter15*

Description

A dataset containing affairs,gender,age...10 variables of 601 objects.

Usage

```
fair
```

Format

A data frame with 601 rows and 10 variables:

ID the ID of objects

affairs times of affairs,0:no;7:4~10times;12:more

gender 1:male;0:female

age the age of objects

yearsmarried married time

children 0:no;1:yes

religiousness 1~5:the degree of being religious upward

education 9:junior high school;12:high school;14,16,17,18,20:phd or others

occupation 1~7:by Hollingshead

rating 1~5:evaluation on marriage myself,from quite sad to very happy

Source

http://farimodel.econ.yale.edu/rayfair/worksd.html

---

fiscal_revenue	*Dataset of national revenue in chapter11*

Description

A dataset containing t,industry,agriculture...8 variables of 15 objects

Usage

fiscal_revenue

Format

a dataframe with 15 rows and 8 variables

t year

industry industry output

agriculture agriculture output

construction construction output

consumption social commodities value

pop total population

disaster disaster area

revenue national revenue

GQtest	*Function of 'Goldfeld-Quandt test' for different variance*

Description

Function of 'Goldfeld-Quandt test' for different variance

Usage

```
GQtest(x, y)
```

Arguments

x, y   Numeric vectors

Value

The outputs are P value and F value

Examples

```
data("plantarea_outputvalue")
GQtest(plantarea_outputvalue$plant_area,plantarea_outputvalue$output_value)
```

grade	*Dataset of grade in chapter13*

Description

A dataset containing GRADE,GPA,TUCE and PSI 4 variables of 32 objects

Usage

```
grade
```

Format

a dataframe with 32 rows and 4 variables

GRADE 1:grade upward;0:others

GPA average grade

TUCE economic grade in history

PSI 1:guided by new teaching method';0:others

---

graduate_apply   *Dataset of graduate application in chapter14*

Description

A dataset containing apply,pared,public and gpa of 400 objects

Usage

```
graduate_apply
```

Format

a data frame with 4 variables of 400 objects

apply willingness,0:unlikely;1:somewhat likely;2:very likely

pared virtual variable for at least one parent is graduate,0:no;1:yes

public virtual variable for ungraduate school,0:private;1:public

gpa average grade

Source

<http://www.ats.ucla.edu/stat/data/ologit.dta>

---

grainproduct   *Dataset of grain production in chapter12*

Description

A dataset containing year,Y,L,M,K 5 variables of 27 objects

Usage

```
grainproduct
```

Format

a dataframe with 27 rows and 5 variables

year time

Y output

L labor population

M area

K chemical fertilizer

---

happy	*Dataset of happiness in chapter14*

Description

A dataset containing happy,money,sex,love,work of 24 objects

Usage

```
happy
```

Format

a dataframe with 24 rows and 5 variables

happy 1~10:the degree of happiness,and 10 is the happinest

money family income,1000$ per unit

sex lover,1:yes;0:no

love emotional index,1:lonely;2:stable relationship;3:strong relationship

work work index,1:no work;3:ok;5:work well

---

heart_maxrate	*Dataset of age and maxrate in chapter10*

Description

A dataset containing Age and MaxRate of 15objects

Usage

```
heart_maxrate
```

Format

a dataframe with 2 variables and 15 objects

Age Age

MaxRate maxrate

---

height_weight	*Dataset of students'height and weight in chapter6*

Description

A data set containing name,sex,age,height and weight 5 variables of 19 objects

Usage

```
height_weight
```

Format

a dataframe with 19 rows and 5 variables

name students' name

sex F:female;M:male

age students' age

height students' height

weight students' weight

---

importS_GDP	*Dataset of china's imports demand and total output 1985s-2003s in chapter11*

Description

A dataset containing year,actual_imports and actual_GDP 3 variables of 19 objects

Usage

```
importS_GDP
```

Format

a dataframe with 19 rows and 3 variables

year year

actual_imports actual_imports

actual_GDP actual_GDP

---

income_pay	*Dataset of income and expenditure in chapter10*

Description

A dataset containing income and expenditure of 99 objects

Usage

```
income_pay
```

Format

a dataframe with 99 rows and income,expenditure 2 variables

income income per month

expenditure expenditure per month

---

income_pay_village	*Dataset of average income and consumption of village 1985s-2003s in chapter11*

Description

A dataset containing t,income,expend and cpi 4 variables of 19 objects

Usage

```
income_pay_village
```

Format

a dataframe with 19 rows and 4 variables

t year

income average pure income per year

expend average expenditure per year

cpi consumption price index

investment	*Dataset of investment in chapter10*

Description

A dataset containing year,total_investment,last_value and last_capital 4 variables of 19 objects

Usage

```
investment
```

Format

a dataframe with 19 rows and 4 variables

year year

total_investment total investment current period

last_value last period stock value

last_capital last period capital stock

listed_company	*Dataset of listed company in chapter13*

Description

A dataset containing ARA,ASSET,ATO,ROA,GROWH,LEV,SHARE and ST 8 variables of 30 objects

Usage

```
listed_company
```

Format

a dataframe with 30 rows and 8 variables

ARA the ratio of receivabls over total asset

ASSET log of total asset

ATO turnover rate of asset

ROA interest rate

GROWH growth rate of sales revenue

LEV leverage factor

SHARE share of the first majority shareholder

ST 1:being treated;0: not

---

math_stat *Dataset of students' math and stat score in chapter6*

Description

A dataset containing math and stat 2 variables of 24 objects

Usage

```
math_stat
```

Format

a dataframe with 24 rows and 2 variables

math students' math score

stat students' stat score

---

median_test *Function of 'median test',and similar to the rank.*

Description

Function of 'median test',and similar to the rank.

Usage

```
median_test(x, median = NA)
```

Arguments

x	Numeric vector
median	Comparasion object

Value

The output are the numbers of positive and negative,and P value.

Examples

```
x <- c(21240,4632,22836,5484,5052,5064,6972,7596,14760,15012,18720,9480,4728,67200,52788)
median_test(x, median = 5080)
wilcox.test(x, mu = 5080)
```

| medicion | *Dataset of population and medicion institutions in sichuan province in 2000s in chapter10* |

Description

A dataset containing region,pop and institutions 3 variables of 16 objects

Usage

```
medicion
```

Format

a dataframe with 16 rows and 3 variables

region several regions in sichuan province

pop population(10 thousand)

institutions number of medicion institutions

| plantarea_outputvalue | *Dataset of agritural plant output and area in chapter11* |

Description

A dataset containing plant_area and output_value 2 variables of 29 objects

Usage

```
plantarea_outputvalue
```

Format

a dataframe with 29 rows and 2 variables

plant_area area for plant

output_value output

---

plantgrowth *Dataset of growth of plant in chapter8*

Description

A dataset containing weight and group 2 variables of 30 objects

Usage

```
plantgrowth
```

Format

a dataframe with 30 rows and 2 variables

weight weight

group ctrl,trt1,trt2

---

plot3D *Function of 'Plot Three Dimensional Graph'*

Description

Function of 'Plot Three Dimensional Graph'

Usage

```
plot3D(x, y, f)
```

Arguments

x, y	Numeric vectors
f	A function

Value

a three dimensional graph

Examples

```
x <- seq(-10, 10, length = 30)
y <- x
f <- function(x, y){r <- sqrt(x ^ 2 + y ^ 2);10 * sin(r) / r}
plot3D(x, y, f)
```

---

proptest   *Function of 'Proportion Hypothesis Test'*

Description

Function of 'Proportion Hypothesis Test'

Usage

```
proptest(x, n, p, alternative)
```

Arguments

x	The number of objects of interest
n	The total number
p	The comparision object
alternative	Twoside or less test

Value

The output are the value of the statistic(u) and the p value

Examples

```
proptest(45, 100, 0.5, alternative = 'twoside')
proptest(450, 1000, 0.5, alternative = 'twoside')

#two sample test
prop.test(c(45, 56), c(45 + 35, 56 + 47))
```

RDA   *Datasets, functions and examples from the book:R Data Analysis-Methods and application.*

Description

Datasets, functions and examples from the book:R Data Analysis-Methods and application.

EDA

Function of 'Exploratory Data Analysis'.

u_test

Function of testing for 'Mean of Norm Distribution' where variance is known.

var_test

Function of 'Variance Test' for single sample.

GQtest

Function of 'Goldfeld-Quandt test' for different variance.

median_test

Function of 'median test',and similar to the rank.

proptest

Function of 'Proportion Hypothesis Test'.

plot3D

Function of 'Plot Three Dimensional Graph'.

---

revenue_tax *Dataset of national revenue and tax 1985s-2001s in chapter6*

Description

A dataset containing year revenue and tax 3 variables of 12 objects

Usage

revenue_tax

Format

a dataframe with 12 rows and 3 variables

year time

revenue national revenue(per billion)

tax national tax

---

ship *Dataset of five kinds of ship accidents in chapter15*

Description

A dataset containing Type,TA,TE...15 variables of 40 objects,and 34 objects are available in fact.

Usage

ship

Format

A data frame with 40 rows and 15 variables:

obs objects

Type ship type

TA virtual variable for ship type

TB virtual variable for ship type

TC virtual variable for ship type

TD virtual variable for ship type

TE virtual variable for ship type

T6064 virtual variable during manufacturing

T6569 virtual variable during manufacturing

T7074 virtual variable during manufacturing

T7579 virtual variable during manufacturing

O6074 virtual variable during running

O7579 virtual variable during running

Mon measure of service times

Acc times of accidents

---

tax | *Dataset of national tax revenue 1978s-2012s in chapter10*

Description

A dataset containing time,tax,GDP,expand and CPI 5 variables of 35 objects

Usage

```
tax
```

Format

a dataframe with 35 rows and 5 variables

time time

tax tax revenue

GDP GDP

expand expanditure

CPI consumption price index

---

travel *Dataset of national travel revenue 1994s-2003s in chapter10*

Description

A dataset containing year,travelincome,population,income.town... 7 variables of 10 objects

Usage

```
travel
```

Format

a dataframe with 10 rows and 7 variables

year year

travelincome travelincome

population travel population

income.town average travel expenditure in town

income_village average travel expenditure in village

distance_road distance on road

distance_railway distance on railway

---

treament *Dataset of treament, a kind of censored data, in chapter15*

Description

A dataset containing treament,housesize,income2,outcome2,chronicle and Hukou of 14 objects.

Usage

```
treament
```

Format

A data frame with 14 rows and 6 variables:

treament the number of treament in hospital

housesize the population of the house

income2 income

outcome2 outcome

chronicle ever has treament history

hukou ever has hukou

u_test *Function of testing for 'Mean of Norm Distribution' where variance is known*

Description

Function of testing for 'Mean of Norm Distribution' where variance is known

Usage

```
u_test(data, mu, thegma, alternative = "twoside")
```

Arguments

data	Numeric vector
mu	comparison object
thegma	Numeric value given
alternative	twoside or less

Value

mean(u)value and p value

Examples

```
b <- c(22, 24, 21, 24, 23, 24, 23, 22, 21, 25)
u_test(b, 25, 2.4, alternative = 'twoside')
```

var_test *Function of 'Variance Test' for single sample*

Description

Function of 'Variance Test' for single sample

Usage

```
var_test(x, sigma2)
```

Arguments

x	Numeric vector
sigma2	Numeric value

Value

The output are variance,chisq2 value,df and P value

Examples

```
x <- rnorm(20, 500, 20)
var_test(x, 400)
```

Index