Package 'npsm'

May 26, 2024

Type Package

Title Nonparametric Statistical Methods

Version 2.0.0

Date 2024-05-25

Author John Kloke [aut, cre],
Joseph McKean [aut]

Maintainer John Kloke <johndkloke@gmail.com>

Description Accompanies the book ``Nonparametric Statistical Methods Using R, 2nd Edition'' by Kloke and McKean (2024, ISBN:9780367651350). Includes methods, datasets, and random number generation useful for the study of robust and/or nonparametric statistics. Emphasizes classical nonparametric methods for a variety of designs --- especially one-sample and two-sample problems. Includes methods for general scores, including estimation and testing for the two-sample location problem as well as Hogg's adaptive method.

License GPL (>= 2)

LazyLoad yes

LazyData yes

URL https://github.com/kloke/npsm, https://github.com/kloke/book

Depends R (>= 3.5.0), Rfit

Imports methods, class, plyr

Suggests boot, survival, sm, HSAUR2, remotes, profileR, car, dplyr, tree

NeedsCompilation no

Repository CRAN

Date/Publication 2024-05-26 08:00:04 UTC

R topics documented:

---

acov231 *Analysis of Covariance Example for a two by three two-way design*

Description

This a simulated data set which is used as an example of analysis of covariance. The data frame acov231 contains the data. The responses are in column 1, column 2 contains the levels of factor A, column 3 contains the levels of factor B, and the 4th column contains the covariate. All true parameters (effects) are 0 in this generated data set.

Usage

```
data(acov231)
```

Format

A data frame with 33 observations and 4 variables.

`response` numeric. the response.

`fA` numeric. factor A with 2 levels.

`fB` numeric. factor B with 3 levels.

`covariate` numeric. a covariate.

References

Kloke, J. and McKean J.W. (2014), *Nonparametric Statistical Methods using R*, Boca Raton, FL: Chapman-Hall.

Examples

```
levs = c(2,3)
data = acov231[,1:3]
xcov = matrix(acov231[,4],ncol=1)
temp = kancova(levs,data,xcov)
```

`aligned.test` *Aligned Rank Test*

Description

Aligned rank test for a group/treatment effect after adjusting for covariates.

Usage

```
aligned.test(x, y, g, scores = Rfit::wscores,...)
```

Arguments

`x`	n by p design matrix
`y`	n by 1 response vector
`g`	n by 1 vector denoting group/treatment membership.
`scores`	Which scores should be used for the fit and the test. An object of class scores.
`...`	optional arguments. passed to rfit.

Details

Data are aligned based on the design matrix x using a rank-based fit via rfit.

Value

`statistic`	The value of the test statistic.
`p.value`	The p-value based on a chisq(k-1) distribution where k is the number of groups/treatments.

Author(s)

John Kloke

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

See Also

`rfit`

Examples

```
y<-rt(30,2)
x<-runif(30)
g<-rep(1:3,each=10)
aligned.test(x,y,g)
```

baseball_players1000 *Career Information for a Random Sample of 1000 Baseball Players*

Description

Demographics and position information on 1000 randomly selected baseball players who debuted after 1945.

Usage

```
data("baseball_players1000")
```

Format

A data frame with 1000 observations on the following 28 variables.

`playerID` a character vector
`birthYear` a numeric vector
`birthMonth` a numeric vector
`birthDay` a numeric vector
`birthCountry` a character vector
`birthState` a character vector
`nameFirst` a character vector
`nameLast` a character vector
`weight` a numeric vector
`height` a numeric vector
`bats` a character vector
`throws` a character vector
`debutYear` a numeric vector
`G_all` a numeric vector
`G_p` a numeric vector
`G_c` a numeric vector
`G_1b` a numeric vector
`G_2b` a numeric vector
`G_3b` a numeric vector
`G_ss` a numeric vector
`G_lf` a numeric vector
`G_cf` a numeric vector
`G_rf` a numeric vector
`G_of` a numeric vector
`G_dh` a numeric vector
`G_ph` a numeric vector
`G_pr` a numeric vector
`pitcher` a logical vector

Details

A random subset of baseball players who debuted after 1945 and played in at least 160 games. Includes information on birth (date and location); height (inches) and weight (pounds); whether they bat left (L), right (R), or switch (B); and games played at each postion. The variable pitcher is a derived variable based on if the majority of games were played as a pitcher (i.e.; G_pr/G_all > 0.5).

Source

https://github.com/chadwickbureau/baseballdatabank

References

https://github.com/chadwickbureau/baseballdatabank/blob/master/readme2014.txt

Examples

```
data(baseball_players1000)
hist(baseball_players1000$weight,xlab="Weight (lbs)",
     probability=TRUE, ylim=c(0,0.02),
     main="Histogram of Weight for 1000 Baseball Players")
lines(density(baseball_players1000$weight,na.rm=TRUE))
```

---

bb2010 *Batting statistics for the 2010 baseball season.*

Description

Batting (average, home runs, RBIs) statistics for 2010 full time players. By full time we mean that the batter had at least 450 official at bats during the season.

Usage

```
data(bb2010)
```

Format

A data frame with 122 observations on the following 3 variables.

`ave` batting average

`hr` home runs

`rbi` runs batted in

Source

baseballguru.com

Examples

```
plot(hr~ave,data=bb2010)
```

`blood.plasma` *Blood plasma measurements related to total triglyceride level*

Description

Data table from Table 9.11 of Hollander and Wolfe (1999). The data consists of triglyceride levels on 13 patients. Two factors, each at two levels, were recorded: Sex and Obesity. The concomitant variables are chylomicrons, age, and three lipid variables (very low-density lipoproteins (VLDL), low-density lipoproteins (LDL), and high-density lipoproteins (HDL)).

Usage

```
data(blood.plasma)
```

Format

A data frame with 13 observations on 8 variables.

`Total` Triglyceride level, response

`Sex` Sex, 2 levels

`Obese` Obesity, 2 levels

`Chylo` Chylomicrons, covariate

`VLDL` Very low density, lipids, covariate

`LDL` Low density, lipids, covariate

`HDL` High density, lipids, covariate

`Age` Age

Source

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

References

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

Examples

```
data(blood.plasma)
plot(Total~Age,data=blood.plasma)
boxplot(Total~Obese,data=blood.plasma)
```

brewers1982 *Basic Summaries of Boxscores for the Milwaukee Brewers 1982 Season*

Description

Basic Summaries of Boxscores for the Major League Baseball team Milwaukee (WI) Brewers 1982 Season. The Brewers won the American League championship that year. Brewer, Robin Yount won the Most Valueable Player (MVP) award. #Robin Yount. MVP.

Usage

```
data("brewers1982")
```

Format

A data frame with 163 observations on the following 8 variables.

`Date` a character vector

`Opp` a character vector

`R` a numeric vector

`RA` a numeric vector

`Time` a character vector

`Attendance` a numeric vector

`home` a logical vector

`win` a logical vector

Examples

```
data(brewers1982)
# proportion of wins for a given number of runs scored
pwin <- with(brewers1982,tapply(win,R,mean))
pwin
# graphical display of the above
plot(names(pwin),pwin,xlab='Runs', ylab='Proportion of Wins',main='Brewers 1982')
```

cancertrt *Survival time based on two treatments*

Description

Survival times (in days) for undergoing standard treatment (S) and a new treatment (N).

Usage

```
data("cancertrt")
```

Format

A data frame with 17 observations on the following 3 variables.

`time` Survival time in days

`event` Indicator for event

`trt` a factor with levels `N` `S`

References

Higgins (2004), *Introduction to Modern Nonparametric Statistics*, Pacific Grove, CA:Brooks/Cole–Thomson Learning

Examples

```
data(cancertrt)
with(cancertrt,gehan.test(time,event,trt))
```

centerx *Center Matrix*

Description

Centers a matrix.

Usage

```
centerx(x)
```

Arguments

`x`	a matrix

Details

Returns a centered matrix, i.e., each column of the matrix is replaced by deviations from its column mean.

Value

The centered matrix.

Author(s)

John Kloke, Joseph McKean

See Also

scale

Examples

```
x <- cbind(seq(1,5,length=5),seq(10,20,length=5))
xc <- centerx(x)
apply(xc,1,mean)
```

---

cloud *Cloud Dewpoint*

Description

A regression example with response cloud point of a liquid and predictor the percent of Iodine 8 added to the liquid; see Chapter 3 of Hettmansperger and McKean (2011) or Exercise 4.9.10 of Kloke and McKean (2014)/Exercise 4.7.7 of Kloke and McKean (2024).

Usage

```
data(cloud)
```

Format

Nineteen observations on two variables.

`cloud.point` Cloud point of the liquid

`I8` Percent Iodine 8 added

Source

Draper, N.R. and Smith, H. (1966), *Applied Regression Analysis*, New York: John Wiley and Sons.

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods Using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods Using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
rfit(cloud.point ~ I8,data=cloud)
```

`cor.boot.ci` *Confidence interval for a correlation based on a bootstrap.*

Description

Returns a bootstrap confidence interval for any of the correlations available in the base R `cor` function.

Usage

```
cor.boot.ci(x, y, method = "spearman", conf = 0.95, nbs = 3000)
```

Arguments

`x`	n by 1 vector
`y`	n by 1 vector
`method`	Which correlation to use. Argument passed to `cor`.
`conf`	Confidence level.
`nbs`	number of bootstrap samples to base CI on.

Details

Obtains a percentile bootstrap confidence interval. The bootstrap samples are obtained via the function `boot`.

Value

A confidence interval.

Author(s)

John Kloke, Joseph McKean

See Also

See Also as `cor`

Examples

```
library(boot)
with(bb2010,cor.boot.ci(ave,hr))
```

---

energy *Energy as a Function of temperature difference.*

Description

A regression example with response energy output in watts and the predictor temperature difference in degrees Kevin; see Devore (2012) and Exercise 4.9.11 of Kloke and McKean (2014)/Exercise 4.7.8 of Kloke and McKean (2024).

Usage

```
data(energy)
```

Format

Twenty-four observations on two variables.

`output` Energy output in watts

`temp.diff` Temperature difference in K

Source

Devore, J. (2012), *Probaility and statistics for engineering and the sciences, 8th ed.*, Boston: Brooks/Cole.

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
rfit(output ~ temp.diff,data=energy)
```

`firstbase` *Rounding First Base.*

Description

The amount of time it took 22 baseball players to round first base for each of three methods of rounding.

Usage

```
data(firstbase)
```

Format

A data frame with 22 observations on the following 3 variables.

`round.out` Time when using round out method.

`narrow.angle` Time when using narrow angle method.

`wide.angle` Time when using wide angle method.

Details

Rounding methods are illustrated in Figure 7.1 of Hollander and Wolfe (1999).

Source

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

References

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

`fk.test` *Two-sample Fligner-Kileen test for homogeneous scales.*

Description

Returns the Fligner-Kileen test for homogeneous scales for two-samples. Also estimates of ratio of scales based on the logs of folded median-aligned samples and a corresponding confidence interval is computed. fk.test computes the value of the statistic based on squared-normal scores following the optimal (for normal errors) such test described in Section 2.10 of Hettmansperger and McKean (2011). Hence, it will differ from the core R routine fligner.test; see the discussion in Section 3.3 of Kloke and McKean (2014)/Section 3.5 of Kloke and McKean (2024).

Usage

```
fk.test(x,y,alternative = c("two.sided", "less", "greater"),conf.level = 0.95)
```

Arguments

x	vector of first sample responses
y	vector of second sample responses
alternative	alternative indicator for hypotheses
conf.level	confidence coefficient for the returned confidence intervals

Details

Returns the Fligner-Kileen test for the two-sample scale problem.

Value

statistic	chi-squared test statistic
p.value	p-value of the test
estimate	vector of estimates of ratio of scales
conf.int	table of confidence intervals

Author(s)

John Kloke, Joseph McKean

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

See Also

fkk.test

Examples

```
x<-rnorm(18)
y<-rnorm(22)*3
fk.test(x,y)
```

fkk.test *k-Sample version of the Fligner-Kileen test for homogeneous scales.*

Description

Returns the Fligner-Kileen test for homogeneous scales for k-samples. Also estimates of ratio of scales based on the logs of folded median-aligned samples and a corresponding confidence interval is computed. The first level (sample) is referenced. See the discussion in Section 5.7 of Kloke and McKean (2014)/Section 5.8 of Kloke and McKean (2024).

Usage

```
fkk.test(y,ind,conf.level = 0.95)
```

Arguments

y	vector of responses
ind	vector of corresponding levels
conf.level	confidence coefficient for the returned confidence intervals

Details

Returns the Fligner-Kileen test for the k-sample scale problem.

Value

statistic	chi-squared test statistic
p.value	p-value of the test
estimate	vector of estimates of ratio of scales
conf.int	table of confidence intervals
cwts	vector of weights based on the estimates difference in scales

Author(s)

John Kloke, Joseph McKean

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

See Also

fk.test

Examples

```
y1 <- rnorm(10)
y2 <- rnorm(12)*3
y3 <- rnorm(15)*5
y<-c(y1,y2,y3)
ind<-rep(1:3,times=c(10,12,15))
fkk.test(y,ind)
```

`fp.test` *Placement Test for the Behrens-Fisher problem.*

Description

Returns the test based on placements for the Behrens-Fisher problem. This test was developed by Fligner and Policello (1981); see, also, Section 2.11 of Hettmansperger and McKean (2011) and Section 4.4 of Hollander and Wolfe (1999). The version computed by fp.test is discussed in Section 3.4 of Kloke and McKean (2014)/Section 3.6 of Kloke and McKean (2024).

Usage

```
fp.test(x,y,delta0=0,alternative = "two.sided")
```

Arguments

`x`	vector of first sample responses
`y`	vector of second sample responses
`delta0`	null value tested
`alternative`	alternative indicator for hypotheses

Details

Returns the Placement Test for the Behrens-Fisher problem.

Value

`statistic`	chi-squared test statistic
`p.value`	p-value of the test
`numerator`	numerator of test statistic
`denominator`	denominator of test statistic

Author(s)

John Kloke, Joseph McKean

References

Fligner, M.~A. and Policello, G.~E. (1981), Robust rank procedures for the Behrens-Fisher problem, *Journal of the American Statistical Association*, 76, 162–168.

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

Hollander, M. and Wolfe, D.~A. (1999), *Nonparametric statistical methods, 2nd Edition*, New York: John Wiley and Sons.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

`gehan.test` *Gehan generalization the Wilcoxon two-sample test*

Description

Generalization of the Wilcoxon rank sum which allows for censored data.

Usage

```
gehan.test(time, event, trt)
```

Arguments

`time`	Time of event or of censoring
`event`	Indicator variable representing a event occur or not (time is censored)
`trt`	Variable indicating treatment group.

Value

`statistic`	Value of the test statistic
`p.value`	p-value

Author(s)

John Kloke

References

Higgins (2004), *Introduction to Modern Nonparametric Statistics*, Pacific Grove, CA:Brooks/Cole–Thomson Learning

Examples

```
n<-76
y<-rexp(n)
event<-rbinom(n,1,0.7) # about 30%  censored
trt<-sample(c(0,1),n,replace=TRUE)
gehan.test(y,event,trt)
```

---

getxact *Design Function for Robust Analysis of Covariance*

Description

Returns the hetrogeneous slopes design matrix used in ANCOVA. It refereences the first level.

Usage

```
getxact(amat,bmat)
```

Arguments

amat	cell mean design matrix of factor.
bmat	matrix of covariates.

Details

Returns the heterogeneous slopes analysis of covariance matrix.

Value

cmat	heterogeneous slopes analysis of covariance matrix

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

getxact2 *Design Function for Robust Analysis of Covariance*

Description

Returns the hetrogeneous slopes design matrix used in ANCOVA. It refereences the first level. Also, column names are supplied.

Usage

```
getxact2(amat,bmat)
```

Arguments

amat	cell mean design matrix of factor.
bmat	matrix of covariates.

Details

Returns the heterogeneous slopes analysis of covariance matrix.

Value

cmat	heterogeneous slopes analysis of covariance matrix eith columns named

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

hemorrhage *Hemorrhage data from Dupont.*

Description

Hemorrhage data from Dupont.

Usage

```
data(hemorrhage)
```

Format

A data frame with 71 observations on the following 3 variables.

`genotype` a numeric vector

`time` a numeric vector

`recur` a numeric vector

References

Dupont

Examples

```
data(hemorrhage)
## maybe str(hemorrhage) ; plot(hemorrhage) ...
```

`hodges_lehmann.ci` *Hodges-Lehmann type estimation and confidence intervals.*

Description

Hodges-Lehmann type estimation and confidence intervals.

Usage

```
hodges_lehmann.ci(x, y, var.equal = FALSE, conf.level = 0.95, ...)
```

Arguments

`x`	numeric vector.
`y`	numeric vector.
`var.equal`	logical. Assume scales are equal (TRUE) of not (FALSE).
`conf.level`	confidence level to be used for the confidence interval.
`...`	optional arguments. currently unused.

Details

Currently implements 2-sample estimation and confidence intervals based on methods purposed by Hodges and Lehnmann.

Value

`estimate`	parameter point estimate
`stderr`	estimated standard error of point estimate
`conf.int`	estimated confidence interval

Author(s)

John Kloke, Joseph McKean

References

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall.

See Also

wilcox.test

Examples

```
zoo<-c(390,258,298,255,324,240,416,319,225,284)
rh <- c(187,186,179,269,382,264,353 ,38,350,267,229,383,254,302,195, 43,337,390)
hodges_lehmann.ci(zoo,rh)
```

---

hodgkins *Relapse-Free Survival Times for Hodgkin's Disease Patients*

Description

These data are described in Example~11.7 of Hollander and Wolfe (1999). Results from a clinical trial in early Hodgkin's disease. Subjects received one of two treatments: radiation of affected node (AN) or total nodal radiation (TN).

Usage

```
data("hodgkins")
```

Format

A data frame with 49 observations on the following 3 variables.

`time` Survival time

`relapse` Indicator variable for relapse

`trt` treatment: a factor with levels `AN` `TN`

References

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

`hogg.test` *Hogg's Adaptive Test*

Description

Based on selector statistics (Q1 & Q2) one of four score functions is choosen. A rank test and p-value is then calculated based on it.

Usage

```
hogg.test(x, y, ...)
```

Arguments

`x`	n by 1 vector
`y`	m by 1 vector
`...`	additional arguments. currently not used

Value

`statistic`	Value of the test statistic.
`p.value`	p-value based on a normal approximation.
`scores`	Which of the score functions was choosen.

Author(s)

John Kloke, Patrick Kimes

References

Hogg, R. McKean, J, Craig, A (2013) *Introduction to Mathematical Statistics, 7th Ed.* Boston: Pearson.

Examples

```
hogg.test(rt(20,1),rt(22,1)+0.2)
```

HoggsQs *Hogg's Q1 and Q2.*

Description

Q1 is a measure of skewness and Q2 is a measure of tail heaviness.

Usage

```
Q1(z)
```

Arguments

z	n by 1 vector

Details

Used as selector statistics in adaptive schemes. Both Q1 and Q2 are ratios. For Q1, the numerator is upper 5% mean minus the middle 50% mean, while the denominator is difference between the middle 5% mean and the lower 5% mean. For Q2, the numerator is upper 5% mean minus the lower 5% mean, while the denominator is difference between the upper 50% mean and the lower 50% mean. These statistics are not robust.

Value

Returns the calculated ratio as a numeric scalar.

Author(s)

John Kloke

References

Hogg, R. McKean, J, Craig, A (2013) *Introduction to Mathematical Statistics, 7th Ed.* Boston: Pearson.

See Also

hogg.test

huitema496 *Analysis of Covarince Data Set*

Description

A data set presented on Page 496 of huitema (2011). The design is a 2 by 2 with one covariate.

Usage

```
data(huitema496)
```

Format

A 16 by 4 array with the following 4 columns:

- `y` number of novel responses.
- `i` type of reinforcement (2 levels).
- `j` type of program (2 levels).
- `x` covariate, a measure of verbal fluency.

Details

Discussion can be found in both references listed below.

Source

Huitema, B.E. (2011), *The analysis of covariance and alternatives, 2nd ed.*, New York: Wiley.

References

Huitema, B.E. (2011), *The analysis of covariance and alternatives, 2nd ed.*, New York: Wiley.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

Examples

```
huitema496 <- data.frame(huitema496)
fit <- rfit(y~factor(i)+factor(j)+x,data=huitema496)
summary(fit)
```

`insulation` *Insulating Fluid Data*

Description

Study the breakdown time of an electrical insulating fluid subject to seven different levels of voltage stress.

Usage

```
data("insulation")
```

Format

A data frame with 76 observations on the following 2 variables.

`log.stress` log of voltage stress

`log.time` log of failure time

Source

Nelson, W. (1982), *Applied lifetime data analysis*, New York: John Wiley and Sons.

Lawless, J.F. (1982), *Statistical models and methods for lifetime data*, New York: John Wiley and Sons.

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

Examples

```
myscores <- logGFscores
myscores@param <- c(1,5)
fit <- rfit(log.time ~ log.stress,scores=myscores,data=insulation)
summary(fit)
fit$tauhat
```

internal *Internal Functions*

Description

Internal functions not intended for general use. Used in calculation of Hogg's Qs.

Usage

```
lmean(z, p)
```

Arguments

z	n by 1 vector
p	scalar

Value

Returns the calculated value as a numeric scalar.

Author(s)

John Kloke, Joseph McKean

See Also

hogg.test,HoggsQs

jonckheere *Jonckheere's Test for Ordered Alternatives*

Description

Computes Jonckheere's Test for Ordered Alternatives; see Section 5.6 of Kloke and McKean (2014)/Section 5.7 of Kloke and McKean (2024).

Usage

```
jonckheere(y, groups)
```

Arguments

y	vector of responses
groups	vector of associated groups (levels)

Details

Computes Jonckheere's Test for Ordered Alternatives. The main source was downloaded from the site:

smtp.biostat.wustl.edu/sympa/biostat/arc/s-news/2000-10/msg00126.html

Value

Jonckheere	test statistic
ExpJ	null expectation
VarJ	null variance
p	p-value

Author(s)

John Kloke, Joseph McKean

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

smtp.biostat.wustl.edu/sympa/biostat/arc/s-news/2000-10/msg00126.html

Examples

```
 r<-rnorm(30)
 gp<-c(rep(1,10),rep(2,10),rep(3,10))
jonckheere(r,gp)
```

---

kancova *Robust Analysis of Covariance under Heterogeneous Slopes for a k-way layout*

Description

Returns a robust rank-based analysis of covariance for a k-way layout assuming heterogenous slopes; see Section 5.4 of Kloke and McKean (2014)/Sections 5.6 and 7.3 of Kloke and McKean (2024). Currently only wilcoxon scores are used.

Usage

```
kancova(levs,data,xcov,print.table=TRUE)
```

Arguments

levs	vector of levels corresponding to the factors A, B, C, etc.
data	matrix with response in column 1 and level in column 2
xcov	matrix of covariates
print.table	logical indicating a table should be printed

Details

Returns the analysis of covariance table assuming heterogenous slopes for a k-way layout.

Value

tab2	analysis of covariance
fint	rank-based ful model (heterogenous slopes
fithomog	rank-based ful model (homogeneous slopes

Author(s)

John Kloke, Joseph McKean

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
levels <- c(2,2)
y.group <- huitema496[,c('y','i','j')]
xcov <- huitema496['x']
kancova(levels,y.group,xcov)
```

---

kancovarown *routine used in the ANCOVA table obtained by kancova*

Description

routine used in making the display of the ANCOVA table obtained by kancova.

Usage

```
kancovarown(vec)
```

Arguments

vec	vector to be labeled.

Details

Returns the labels.

Value

nm	vector of labels

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

---

knn_cv *Train a k nearest neighbors (knn) classifer via cross validation (cv).*

Description

Train a k nearest neighbors (knn) classifer via cross validation (cv). The number of folds and the set of the number of neihbors to consider may be specified.

Usage

```
knn_cv(xy, k.cv = 5, kvec = seq(1, 47, by = 2))
```

Arguments

xy	Data frame with the data matrix x as the first set of columns and the vector y as the last column.
k.cv	scalar. number of folds to use. default is 5.
kvec	vector. set of neighbors to consider. default is odd integers between 1 and 47 (inclusive).

Value

kvec	set of neighbors considered
error	vector of misclassification error rates corresponding to kvec
k.best	number of neighbors with lowest error rate
k.cv	number of folds to used

Author(s)

John Kloke

References

Hastie, T., Tibshiani, R., and Friedman, J. (2017), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, New York: Springer.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), An Introduction to Statistical Learning with Applications in R, New York: Springer.

Venables, W. N. and Ripley, B. D. (2002) _Modern Applied Statistics with S._ Fourth edition. Springer.

See Also

knn

Examples

```
train_set <- sim_class2[sim_class2$train==1,-1]
set.seed(19180511)
fit_cv <- knn_cv(train_set,k.cv=10)
fit_cv
```

---

latour *Chateau Latour Wine Data*

Description

The response variable is the quality of a vintage based on a scale of 1 to 5 over the years 1961 to 2004. The predictor is end of harvest, days between August 31st and the end of harvest for that year, and the factor of interest is whether or not it rained at harvest time.

Usage

```
data(latour)
```

Format

A data frame with 44 rows and 4 columns.

`year` Year of harvest

`quality` Rating on a scale of 1-5

`end.of.harvest` Days August 31 and the end of harvest

`rain` indicator variable for rain

References

Sheather, SJ (2009), *A Modern Approach to Regression with R*, New York: Springer.

Examples

```
data(latour)
plot(quality~end.of.harvest,pch='',data=latour)
points(quality~end.of.harvest,data=latour[latour$rain==0,],pch=3)
points(quality~end.of.harvest,data=latour[latour$rain==1,],pch=4)
```

---

mood.ci *Mood Median Confidence Interval*

Description

Mood's classical nonparametric method for calculating a difference in population medians.

Usage

```
mood.ci(x, y, var.equal = FALSE, conf.level = 0.95, ...)
```

Arguments

x	n x 1 vector
y	m x 1 vector
var.equal	Logical. Assume scale of the two populations are equal.
conf.level	numeric value. confidence level for the confidence interval.
...	not currently implmented

Value

A vector of length 2 containing the lower and upper endpoints of the confidence interval.

Author(s)

John Kloke, Joseph McKean

References

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall.

See Also

hl.ci,wilcox.test

Examples

```
x <- rt(101,9)
y <- rt(108,9)+0.3
mood.ci(x,y)
```

onecova *Robust Analysis of Covariance under Heterogeneous Slopes*

Description

Returns tests for homogeneous slopes and also assuming homogeneous slopes a test for differences in level. Currently only wilcoxon scores are used.

Usage

```
onecova(levs,data,xcov,print.table=TRUE)
```

Arguments

levs	Number of levels of the one-way design
data	matrix with response in column 1 and level in column 2
xcov	matrix of covariates
print.table	logical indicating a table should be printed

Details

Returns the analysis of covariance table.

Value

tab	analysis of covariance

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

Examples

```
data=latour[,c('quality','rain')]
xcov<-cbind(latour['end.of.harvest'])
onecova(2,data,xcov,print.table=TRUE)
```

onecovaheter *Robust Analysis of Covariance under Heterogeneous Slopes*

Description

Returns a robust rank-based analysis of covariance for a one-way layout assuming heterogenous slopes; see Section 5.4 of Kloke and McKean (2014)/Sections 5.6 and 7.3 of Kloke and McKean (2024). Currently only wilcoxon scores are used.

Usage

```
onecovaheter(levs,data,xcov,print.table=TRUE)
```

Arguments

levs	Number of levels of the one-way design
data	matrix with response in column 1 and level in column 2
xcov	matrix of covariates
print.table	logical indicating a table should be printed

Details

Returns the analysis of covariance table assuming heterogenous slopes.

Value

tab	analysis of covariance
fit	rank-based ful model (heterogenous slopes

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
data=latour[,c('quality','rain')]
xcov<-cbind(latour['end.of.harvest'])
onecovaheter(2,data,xcov,print.table=TRUE)
```

onecovahomog *Robust Analysis of Covariance under Heterogeneous Slopes*

Description

Returns a robust rank-based analysis of covariance for a one-way layout assuming homogeneous slopes; see Section 5.4 of Kloke and McKean (2014)/Sections 5.6 and 7.3 of Kloke and McKean (2024). Currently only wilcoxon scores are used.

Usage

```
onecovahomog(levs,data,xcov,print.table=TRUE)
```

Arguments

levs	Number of levels of the one-way design
data	matrix with response in column 1 and level in column 2
xcov	matrix of covariates
print.table	logical indicating a table should be printed

Details

Returns the analysis of covariance table assuming homogeneous slopes.

Value

tab	analysis of covariance
fit	rank-based ful model (homogeneous slopes

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
data=latour[,c('quality','rain')]
xcov<-cbind(latour['end.of.harvest'])
onecovahomog(2,data,xcov,print.table=TRUE)
```

`place` *Placements.*

Description

Returns the placements of the first vector in terms of the second vector used the R function fp.test; see Section 2.11 of Hettmansperger and McKean (2011) and Section 4.4 of Hollander and Wolfe (1999). The version computed by fp.test is discussed in Section 3.4 of Kloke and McKean (2014)/Section 3.6 of Kloke and McKean (2024).

Usage

```
place(x,y)
```

Arguments

`x`	first vector
`y`	second vector of second sample responses

Details

Returns the Placements for the routine fp.test.

Value

`ic`	vector of placements.

Author(s)

John Kloke, Joseph McKean

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

Hollander, M. and Wolfe, D.~A. (1999), *Nonparametric statistical methods, 2nd Edition*, New York: John Wiley and Sons.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

plank *Plank data*

Description

Abebe et al. (2001) discuss a dataset resulting from a three-way layout for a neurological experiment in which the time required for a mouse to exit a narrow elevated wooden plank is measured. The response is the log of time (in seconds) to exit. Interest lies in assessing the effects of three factors: the Mouse Strain (Tg+, Tg-), the mouse's Gender (female, male), and the mouse's Age (Aged, Middle, Young). The design is a 2 by 2 by 3 factorial design.

Usage

```
data(plank)
```

Format

A data frame with 64 observations on the following 4 variables.

`response` a numeric vector

`strain` a factor with levels `1` `2`

`gender` a factor with levels `1` `2`

`age` a factor with levels `1` `2` `3`

References

Abebe, A., Crimin, K., McKean, J. W., Vidmar, T. J., and Haas, J. V. (2001) "Rank-Based Procedures for Linear Models: Applications to Pharmaceutical Science Data" *Drug Information Journal*,

Examples

```
data(plank)
boxplot(response~strain,data=plank)
raov(response~strain:gender:age,data=plank)
```

plot.knn_cv *plot function for knn_cv*

Description

plots the misclassification error rate versus number of neighbors based on call to knn_cv

Usage

```
## S3 method for class 'knn_cv'
plot(x, ...)
```

Arguments

x	object of class knn_cv.
...	additional arguments. currently not used.

Details

The list x is assumed to have attributes kvec and error representing the number of neighbors and the corresponding misclassification rate, respectively.

Value

No return value, called for side effects of creating plot.

Author(s)

John Kloke

References

Hastie, T., Tibshiani, R., and Friedman, J. (2017), The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, New York: Springer.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), An Introduction to Statistical Learning with Applications in R, New York: Springer.

Venables, W. N. and Ripley, B. D. (2002) _Modern Applied Statistics with S._ Fourth edition. Springer.

See Also

knn_cv

---

poly *A Simulated Polynomial Data Set.*

Description

A simulated polynomial (3rd degree) model discussed in Section 4.7.1 of Kloke and McKean (2014)/4.6.1 of Kloke and McKean (2024).

Usage

```
data(poly)
```

Format

One-hundred observations on two variables.

y response variable

x predictor

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
plot(y ~ x,data=poly)
```

`polydeg` *Degree of Polynomial Determination*

Description

Tests for the degree of a polnomial. This test was suggested by Graybill (1976) and is discussed from a robust point-of-view in Section 4.7.1. of Kloke and McKean (2014)/4.6.1 of Kloke and McKean (2024).

Usage

```
polydeg(y, x, P, alpha = 0.05)
```

Arguments

`y`	vector of responses
`x`	Predictor
`P`	Super degree of polynomial which provides a satisfactory fit
`alpha`	Level of the testing

Details

Returns the degree of the polynomial based on the algorithm.

Value

`deg`	The determined degree
`coll`	Matrix of step information
`fitf`	Fit of the polynomial based on the determoned degreer

References

Graybill, F.A. (1976), *Theory and application of the linear model*, North Scituate, Ma: Duxbury Press.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
x <- 1:20
xc <- x - mean(x)
y<- .2*xc + xc^3 +rt(20,3)*90
plot(y~x)
polydeg(y,xc,6)
```

---

print *Internal print functions*

Description

Internal print functions

Usage

```
## S3 method for class 'hogg.test'
print(x, digits = max(5, .Options$digits - 2), ...)
## S3 method for class 'rank.test'
print(x,...)
## S3 method for class 'fkk.test'
print(x,...)
## S3 method for class 'knn_cv'
print(x,...)
## S3 method for class 'npsm.ci'
print(x, estimate=FALSE,stderr=FALSE,digits = max(5, .Options$digits - 2),...)
```

Arguments

x	Object to be printed.
digits	Number of digits to present. Passed to print function.
...	Additional arguments.
estimate	not currently implemented.
stderr	not currently implemented.

Value

No return value, called for side effects

Author(s)

John Kloke, Joseph McKean

`prostate` *DES for treatment of prostate cancer.*

Description

Under investigation in this clinical trial was the pharmaceutical agent diethylstilbestrol DES; subjects were assigned treatment to 1.0 mg DES (treatment = 2) or to placebo (treatment = 1).

Usage

```
data(prostate)
```

Format

A data frame with 38 observations on the following 8 variables.

`patient` a numeric vector

`treatment` a numeric vector

`time` a numeric vector

`status` a numeric vector

`age` a numeric vector

`shb` a numeric vector

`size` a numeric vector

`index` a numeric vector

Source

http://www.crcpress.com/product/isbn/9781584883258

References

Collett, D. (2003) *Modeling survival data in medical research* CRC press.

Examples

```
data(prostate)
boxplot(size~treatment,data=prostate)
```

qhic *qhic*

Description

A regression example with response yearly upkeep of a home and the predictor value of home; see Bowerman et al. (2005) and Exercise 4.9.8 of Kloke and McKean (2014)/Exercise 7.6.2 of Kloke and McKean (2024).

Usage

```
data(qhic)
```

Format

Forty observations on two variables.

`upkeep` annual upkeep expenditure of home (y)

`value` value of the home (x)

References

Bowerman, B.L., O'Connell, R.T., and Koehler, A.B. (2005), *Forecasting, time series, and regression: An applied approach*, Australia: Thomson.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
plot(upkeep~value,data=qhic,xlab='Value (in $1000s)',ylab='Annual upkeep (in $10s)')
```

---

quail2 *Quail from a two-factor experiment.*

Description

Two sample quail data.

Usage

```
data(quail2)
```

Format

A data frame with 30 observations on the following 2 variables.

`treat` indicator variable for treatment

`ldl` ldl measurement

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

McKean J.W., Vidmar, T.J., and Sievers, G.L. (1989), A robust two stage multiple comparison procedure with application to a random drug screen, *Biometrics*, 45, 1281–1297.

Examples

```
data(quail2)
boxplot(ldl~treat,data=quail2)
```

---

rank.test *General scores rank test for two sample problem*

Description

A generalization of the Wilcoxon rank-sum test where a score function is applied to the ranks. Any scores from Rfit can be used as well as user defined. Default is to perform a Wilcoxon analysis.

Usage

```
rank.test(x, y, alternative = "two.sided", scores = Rfit::wscores,
  conf.int = FALSE, conf.level = 0.95)
```

Arguments

`x`	m x 1 vector
`y`	n x 1 vector
`alternative`	one of 'two.sided', 'less', or 'greater'
`scores`	an object of class scores
`conf.int`	logical indicating if a confidence interval should be estimated
`conf.level`	desired level of confidence for interval

Details

Test is based on T = sum_i a(R(y_i)) where R is the rank based on the combined sample and a(t) = varphi(t/(N+1)). Confidence interval, if requested, is based on call to Rfit.

Value

statistic	Standardized value of test statistics
Sphi	Test statistic
p.value	p-value
conf.int	confidence interval for shift in location
estimate	point estimate for shift in location

Author(s)

John Kloke, Joseph McKean

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

See Also

wilcox.test

Examples

```
rank.test(rt(20,1),rt(22,1)+0.2)
```

---

rcn *random contaminated normal deviates*

Description

Generate a random sample from a contaminated normal distribution.

Usage

```
rcn(n, eps, sigmac)
rcn_5_5(n)
```

Arguments

n	sample size
eps	proportion of proportion of contamination
sigmac	standard devation of contaiminated component

Details

With probability (1-eps) a deviates are drawn from a standard normal distribution. With probability eps deviates are drawn from a normal distribution with mean 0 and standard devation sigmac rcn_5_5 is a special case where eps=0.05 and sigma=5.

Value

n x 1 numeric vector containing the random deviates.

Author(s)

John Kloke, Joseph McKean

References

Hogg, R. McKean, J, Craig, A (2013) *Introduction to Mathematical Statistics, 7th Ed.* Boston: Pearson.

See Also

rnorm

Examples

```
qqnorm(rcn(100,.25,10))

set.seed(101); rcn(10,0.05,5)
set.seed(101); rcn_5_5(10)
```

---

rcnx *Fat-Finger Error Contaminated Normal Deviates*

Description

Generate random data from a contaminated normal distribution where the contaimation is a multiplicative factor. As, for example, in cases of data recorded in incorrect units or incorrect decimal point.

Usage

```
rcnx100(n,eps=0.001,x=100,mu=0,sigma=1,...)
rcnx(...)
rcnx_01_100(n)
```

Arguments

n	sample size to be drawn.
eps	amount (probability) of contaminated observations
x	multiplier for the contaminated observations
mu	mean of uncontaminated samples
sigma	standard deviation of uncontaminated samples
...	optional arguments.

Details

Samples are drawn from a random normal distribution with mean mu and standard deviations. A fraction of the observations (eps) are multiplied by the factor x. rcnx is an alias for rcnx100. rcnx_01_100 is a special case where the observations are drawn from a standard normal distribution (i.e., mu=0 and sigma=1 — the defaults in rcnx100) and eps and x are specified as 0.01 and 100, respectively.

Value

Numeric vector of length n is returned.

Author(s)

John Kloke

References

https://en.wikipedia.org/wiki/Fat-finger_error

See Also

rcn

Examples

```
set.seed(101); x1 <- rcnx100(10)
set.seed(101); x2 <- rcnx(10)
set.seed(101); x3 <- rcnx_01_100(10)

qqnorm(rcnx(10000,eps=0.005,x=10))
qqnorm(rcnx(1000,eps=0.05,x=1/100))
```

---

rlaplace *Random Laplace.*

Description

Random generation for the Laplace (double exponential) data with location 0 and scale 1.

Usage

```
rlaplace(n)
```

Arguments

n	scalar. number of random draws.

Details

A Laplace or double expoential distribution has heavier tails than a normal distribution and so a sample will tend to have additional outliers.

Value

A vector of length n is returned containing the random data.

Author(s)

John Kloke, Joseph McKean

References

Hogg, Robert V.; McKean, Joseph; and Craig, Allen T., "Introduction to Mathematical Statistics (6th Edition)" (2005).

Examples

```
x <- rlaplace(100)
qqnorm(x)
```

`rs` *Simulated Regression Model*

Description

A simulated regression model with one response and one predictor. It is discussed in Exercise 6.5.6 of Kloke and McKean (2014)/Exercise 8.11.23 of Kloke and McKean (2024).

Usage

```
data(rs)
```

Format

Fifty observations on two variables.

`y` simulated response

`x` simulated predictor

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
rfit(y ~ x,data=rs)
```

SCUD *Cyclone Data*

Description

A data set discussed in Hollander and Wolfe (1999) and Exercise 5.8.9 of Kloke and McKean (2014)/Exercise 5.9.15 of Kloke and McKean (2024). It contains part of a study on the effects of cloud seeding of cyclones.

Usage

```
data(SCUD)
```

Format

Twenty-one observations on three variables.

`trt` treatment indicator (1) is Seeded and (2) is control

`M` predictor M, the geostrophic meridional circulation index

`RI` measure of precipitation

References

Hollander, M. and Wolfe, D.A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
plot(RI ~ M,data=SCUD)
```

seinfeld *Seinfeld — the sitcom — viewership counts by episode*

Description

Counts of viewers for 9 seasons of Seinfeld

Usage

```
data("seinfeld")
```

Format

A data frame with 180 observations on the following 4 variables.

`episodeNumberOverall` a numeric vector

`season` a numeric vector

`episodeNumberSeason` a numeric vector

`viewers` a numeric vector

Source

Wikipedia https://en.wikipedia.org/wiki/List_of_Seinfeld_episodes (date unknown).

Examples

```
data(seinfeld)
#Comparison boxplots of views versus season
boxplot(viewers~season,data=seinfeld,ylab='Number of Viewers (in millions)',xlab='Season')

# Normal q-q plots for selected seasons.
oldpar_mfrow <- par()$mfrow
par(mfrow=c(2,2))
seasons2display <- c(4,5,6,9)
for( s in seasons2display) {
  v <- seinfeld[seinfeld$season==s,'viewers']
  qqnorm(v,main=paste("Season",s))
  abline(a=median(v),b=mad(v))
}
par(mfrow=oldpar_mfrow)

# Normal q-q plots for selected seasons
# using centered and scaled residuals.
oldpar_mfrow <- par()$mfrow
par(mfrow=c(2,2))
seasons2display <- c(4,5,6,9)
for( s in seasons2display) {
  v0 <- seinfeld[seinfeld$season==s,'viewers']
  v1 <- (v0 - median(v0))/mad(v0)
  qqnorm(v1,main=paste("Season",s))
  abline(a=0,b=1)
}
par(mfrow=oldpar_mfrow)
```

`sievers` *Doksum and Sievers rat data*

Description

Doksum and Sievers (1976) describe an experiment involving the effect of ozone on weight gain of rats. The experimental group consisted of 22 rats which were placed in an ozone environment for seven days, while the control group contained 21 rats which were placed in an ozone-free environment for the same amount of time. The response was the weight gain in a rat over the time period.

Usage

```
data(sievers)
```

Format

A data frame with 45 observations on the following 2 variables.

`group` indicator variable for treatment

`weight.gain` response variable of weight gain

References

Hettmansperger, T.P. and McKean J.W. (2011), *Robust Nonparametric Statistical Methods, 2nd ed.*, New York: Chapman-Hall.

Doksum, K. A. and Sievers, G. L. (1976), Plotting with confidence: Graphical comparisons of two populations, *Biometrika*, 63, 421-434.

Examples

```
data(sievers)
boxplot(weight.gain~group,data=sievers)
```

---

signtest_pvalue *p-value for a one sample sign test*

Description

p-value for a one sample sign test based on the binomial distribution.

Usage

```
signtest_pvalue(x, alternative = "two.sided", theta0 = 0, ...)
```

Arguments

`x`	number vector.
`alternative`	type of alternative hypothesis
`theta0`	null value of the parameter
`...`	optional arguments. currently ignored.

Details

Returns p-value using the binomial distribution.

Value

a numeric scalar — the p-value — is returned

Author(s)

John Kloke, Joseph McKean

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall.

Examples

```
x <- round(rt(19,9) + 2,1)
signtest_pvalue(x,alternative='greater')
S <- sum(x > 0)
M <- sum(x != 0)
1-pbinom(S-1,M,0.5)
x <- round(rt(19,9) + 0,1)
signtest_pvalue(x)
S <- sum(x > 0)
M <- sum(x != 0)
2*min(pbinom(S,M,0.5), 1-pbinom(S-1,M,0.5))
```

---

simon *Simon (the memory game) dataset*

Description

An experiment in which the members of two groups of students each played the game Simon twice.

Usage

```
data("simon")
```

Format

A data frame with 31 observations on the following 3 variables.

`game1` score on first trial

`game2` score on second trial

`class` group variable

Details

Demonstrates the concept of regression toward the mean. Simulated data to represent a realistic realization of the experiment. See Problem 4.9.20 of Kloke and McKean (2014)/Problem 4.7.17 of Kloke and McKean (2024).

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistcal methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
data(simon)
plot(game2~game1,data=simon)
rfit(game2~game1,data=simon)
```

---

sim_class2 — *A simulated classification example with two variables and two classes (labels).*

Description

A simulated classification example with two variables and two classes (labels).

Usage

```
data("sim_class2")
```

Format

A data frame with 1000 observations on the following 4 variables.

`train` an indicator for training and test sets

`x1` an explantory variable

`x2` an explantory variable

`y` response variable - a factor with levels `0` `1`

Details

Random points in the x1,x2 plane were generated. Class labels based on location relative to two circles in the x1,x2 plane with some random variation in the labels simulated.

Examples

```
data(sim_class2)
dim(sim_class2)

train_set <- sim_class2[sim_class2$train==1,]
dim(train_set)

with(train_set,plot(x1,x2,main='Training Set',cex=0.625))
with(train_set,points(x1,x2,main='Training Set',pch=20,col=y,cex=0.625))
```

---

sincos *Sine Cosine Model*

Description

Simulated dataset

Usage

```
data("sincos")
```

Format

A data frame with 197 observations on the following 2 variables.

`x` independent variable

`y` dependent variable

Details

The data were generated using `x <- seq(1,50,by=.25) ; y <- 5*sin(3*x) + 6*cos(x/4)+rnorm(length(x),0,10)`

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall.

Examples

```
data(sincos)
plot(y~x,sincos)

### code to create Figure 4.9 of Kloke & McKean 2014 ###
my.sincos<-sincos
my.sincos$y3<-my.sincos$y
my.sincos$y3[137] <- 800
```

```
plot(y3~x,ylim=c(-50,50),data=my.sincos)
fit4 <- loess(y3 ~ x,data=my.sincos)
# lines(fit4$x,fit4$fitted,lty=2)
with(fit4,lines(x,fitted,lty=2))
fit5 <- loess(y3 ~ x,family="symmetric",data=my.sincos)
with(fit5,lines(x,fitted,lty=1))
legend('bottomleft',legend=c('Local Robust Fit','Local LS Fit'),lty=1:2)
title("loess Fits of Sine-Cosine Data")
```

---

speed *Predict top speed based on miles per gallon*

Description

A sample of 82 cars with variables speed and miles per gallon collected.

Usage

```
data("speed")
```

Format

A data frame with 82 observations on the following 2 variables.

`mpg` Miles per gallon

`sp` a numeric vector

Source

Higgins (2003) Introduction to modern nonparmetric statistics.

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistcal methods using R*, Boca Raton, FL: Chapman-Hall.

Examples

```
data(speed)
plot(sp~mpg,data=speed)
rfit(sp~mpg+I(mpg^2),data=speed)
```

turtle *Turtle Data*

Description

A data frame containg measurements of 48 turtles. The first three columns are the Length, Width, and Height measurements of the carapace of the turtle. The fourth column is a categorical variable sex with values of female and male. Data are drawn from Johnson and Wichern (2007).

Usage

```
data(turtle)
```

Format

48 observations on four variables.

Length numeric vector.

Width numeric vector.

Height numeric vector.

sex character vector.

References

Johnson, R.A. and Wichern, D.W. (2007), *Applied Multivariate Statistical Analysis, 6th ed.*, Upper Saddle River, NJ: Pearson.

Examples

```
with(turtle,boxplot(Length~sex))
with(turtle,boxplot(Length~sex,ylab='Length (units)'))
```

vanElteren.test *vanElteren test for stratified analysis*

Description

Performs the vanElteren extension of the Wilcoxon rank sum test for stratified experiments.

Usage

```
vanElteren.test(g, y, b)
```

Arguments

g	n x 1 vector: treatment/group indicator
y	n x 1 vector: responses
b	n x 1 vector: denotes strata

Value

statistic	Value of the test statistic.
p.value	p-value based on a normal approximation.

---

weather *January Weather Data for Kalamazoo*

Description

January weather data for Kalamazoo, MI for the years 1900 to 1995. It is discussed in Example 4.7.4, page 105-106, of Kloke and McKean (2014)/Example 4.6.4, p.177-178, of Kloke and McKean (2024).

Usage

```
data(weather)
```

Format

Ninety-six observations (1900-1995) for twelve weather variables.

avemax avemax

avemin avemin

coldestmax coldestmax

hihest hihest

lowest lowest

maxdayprec maxdayprec

maxdaysnowfall maxdaysnowfall

meantmp meantmp

totalprec totalprec

totalsnow totalsnow

warmest warmest

year year

Source

http://weather-warehouse.com/WeatherHistory/

References

Kloke, J. and McKean, J.W. (2014), *Nonparametric statistical methods using R*, Boca Raton, FL: Chapman-Hall. Kloke, J. and McKean, J.W. (2024), *Nonparametric statistical methods using R, Second Edition*, Boca Raton, FL: Chapman-Hall.

Examples

```
plot(avemax ~ year,data=weather)
```

`wilson.ci` *Wilson (score) confidence interval for a population proportion.*

Description

Wilson (score) confidence interval for a population proportion.

Usage

```
wilson.ci(x, n, conf.level = 0.95)
```

Arguments

`x`	number of events
`n`	number of samples
`conf.level`	confidence level

Details

Uses defintion in Agresti.

Value

`conf.int`	estimated confidence interval

Author(s)

John Kloke, Joseph McKean

References

Agresti (2002), Categorical data analysis, New York: John Wiley & Sons, Inc.

See Also

`prop.test`

Examples

```
n <- 100
x <- rbinom(1,n,0.33)
wilson.ci(n,x)
```

Index