

Package: quantileCI (via r-universe)

September 15, 2024

Title Confidence Intervals for Quantiles

Version 0.1

Description Computes exact and interpolated confidence intervals for population quantiles based on a single independent and identically distributed sample. The package implements different methods for computing the confidence intervals and also provides functionality to check the coverage of the methods.

Imports boot

Depends R (>= 3.3.0), units

Suggests rmarkdown, knitr, testthat

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

URL <https://github.com/hoehleatsu/quantileCI>

VignetteBuilder knitr

Repository <https://mhoehle.r-universe.dev>

RemoteUrl <https://github.com/mhoehle/quantileCI>

RemoteRef HEAD

RemoteSha 0057ad82cac967f6e3741132f87ba8ee0010818b

Contents

flint	2
median_confint_hs	2
qci_coverage_one_sim	4
quantile_confint_boot	5
quantile_confint_exact	6
quantile_confint_nyblom	7
Index	9

flint

*Water Monitoring Sample from Flint, Michigan, 2015***Description**

The data correspond to the original January as well as the revised June version of the 2015 "Lead and Copper Report and Consumer Notice of Lead Result" report. Data are taken from the sources stated below, not the original report. Note that a lead level of zero means that the measurement was below the detection limit.

Usage

```
data(flint)
```

Format

A data frame with 71 rows and 2 variables:

lead The measured lead concentration in the tap water in parts per billion (ppb = mg/L) as now also made explicit using the `units` package.

exclude Logical indicating if the observations was removed by the authorities or not. In total two observations were removed as explained in the references.

References

Langkjær-Bain, R. (2017), The murky tale of Flint's deceptive water data. *Significance*, 14: 16–21. <https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2017.01016.x>.

Quantiles and the Flint water crisis (2017), Wicklin, R, <http://blogs.sas.com/content/iml/2017/05/17/quantiles-flint-water-crisis.html>

Youtube video by Michigan Radio explaining the computation of the quantile <https://www.youtube.com/watch?v=9pql00zr700>

Beware the Argument: The Flint Water Crisis and Quantiles (2016), Höhle M, <https://staff.math.su.se/hoehle/blog/2017/06/18/quantiles.html>

median_confint_hs

*Two-sided confidence interval method for the median by the method of Hettmansperger & Sheather (1991)***Description**

Two-sided confidence interval method for the median by the method of Hettmansperger & Sheather (1991)

Usage

```
median_confint_hs(  
  x,  
  conf.level = 0.95,  
  x_is_sorted = FALSE,  
  interpolate = TRUE  
)
```

Arguments

x	vector of observations
conf.level	A conf.level * 100% confidence interval is computed
x_is_sorted	Boolean (Default: FALSE) to safe sorting x, if it is already sorted. This is merely for speed reasons in situations where it is more efficient to only sort x once.
interpolate	Boolean (Default: TRUE) stating whether to interpolate the order statistics. If no interpolation is selected then this is just the standard exact procedure based on the order statistics. Note: This procedure is conservative (i.e. coverage is usualler larger than the nominal conf.level and hence the interval is actually in general too large).

Details

The interpolation procedure suggested by Hettmansperger and Sheather (1986) for the median is applied to the order statistic.

Value

A vector of length two containing the lower and upper limit of the confidence interval

References

Hettmansperger TP and Sheather SJ (1986), Confidence intervals based on interpolated order statistics, *Statistics and Probability Letters*, 4, p. 75-79.

Examples

```
set.seed(123)  
x <- rnorm(25)  
median_confint_hs(x=x, conf.level=0.95, interpolate=TRUE)
```

qci_coverage_one_sim *Computing the coverage of different confidence interval methods for quantiles by Monte Carlo integration.*

Description

Computing the coverage of different confidence interval methods for quantiles by Monte Carlo integration.

Usage

```
qci_coverage_one_sim(
  qci_fun,
  n,
  rfunc = rnorm,
  qfunc = qnorm,
  p = 0.5,
  conf.level = 0.95,
  ...
)
```

Arguments

qci_fun	Function which given n, p and conf.level computed a set of different confidence intervals. Should return a matrix of dimension 2 x (no. of methods) which contains the lower and upper bound of each confidence interval method.
n	Size of the sample to generate in the simulation
rfunc	Function for generating the samples
qfunc	Quantile function for computing the true quantile
p	The quantile of interest $0 \leq p \leq 1$
conf.level	conf.level * 100% two-sided confidence intervals are computed
...	Additional arguments passed to rfunc and qfunc

Value

A vector of Booleans of length (no. methods) stating if each method contains the true value or not

Examples

```
##Function to compute different methods on same x.
quantile_confints <- function(x, p, conf.level, x_is_sorted=FALSE) {
  if (!x_is_sorted) { x <- sort(x)}

  ##Compute the various confidence intervals as above
  res <- data.frame(
    nyblom_exact=quantileCI::quantile_confint_nyblom(x=x, p=p, conf.level=conf.level,
```

```

                                x_is_sorted=TRUE, interpolate=FALSE),
  nyblom_interp=quantileCI::quantile_confint_nyblom(x=x, p=p, conf.level=conf.level,
                                x_is_sorted=TRUE, interpolate=TRUE),
  boot=quantileCI::quantile_confint_boot(x, p=p, conf.level=conf.level, R=999)
)
if (p == 0.5) {
  res$hs_interp = quantileCI::median_confint_hs(x=x, conf.level=conf.level,
                                x_is_sorted=TRUE, interpolate=TRUE)
}
return(res)
}

## One run of the simulation function
quantileCI::qci_coverage_one_sim(qci_fun=quantile_confints, n=100,p=0.5,conf.level=0.95)

## Several runs, calculate row means to get coverage by sampling
res <- sapply(1L:10L, function(i) {
  quantileCI::qci_coverage_one_sim(qci_fun=quantile_confints, n=100,p=0.5,conf.level=0.95)
})
res

```

quantile_confint_boot *Confidence interval method for a given quantile based on the basic bootstrap and using the percentile method.*

Description

Confidence interval method for a given quantile based on the basic bootstrap and using the percentile method.

Usage

```
quantile_confint_boot(x, p, conf.level = 0.95, R = 999, type = 7)
```

Arguments

x	vector of observations
p	quantile of interest
conf.level	A conf.level * 100% confidence interval is computed
R	number of replications to use in the bootstrap (Default: 999)
type	Type of empirical quantile estimation procedure, @seealso the quantile function.

Details

Basic bootstrap with the confidence interval computed based on the percentile method.

Value

A vector of length two containing the lower and upper limit of the two-sided confidence interval.

Examples

```
set.seed(123)
x <- rnorm(25)
quantile_confint_boot(x=x, p=0.8, conf.level=0.95, R=999)
```

quantile_confint_exact

Standard exact two-sided quantile confidence interval based on the binomial distribution

Description

Standard exact two-sided quantile confidence interval based on the binomial distribution

Usage

```
quantile_confint_exact(
  x,
  p,
  conf.level = 0.95,
  x_is_sorted = FALSE,
  fix_interval = TRUE
)
```

Arguments

x	vector of observations
p	quantile of interest, $0 \leq p \leq 1$
conf.level	A conf.level * 100% confidence interval is computed
x_is_sorted	Boolean (Default: FALSE) to safe sorting x, if it is already sorted. This is merely for speed reasons in situations where it is more efficient to only sort x once.
fix_interval	Boolean (Default: TRUE) For the case with no interpolation, try to extend interval upwards if coverage is too little.

Details

This function is a pure call-through to the Nyblom function with interpolate=FALSE.

Value

A vector of length two containing the lower and upper limit of the confidence interval

References

Nyblom J, Note in interpolated order statistics, Statistics and Probability Letters 14, p. 129-131.

Examples

```
set.seed(123)
x <- rnorm(25)
quantile_confint_exact(x=x, p=0.8, conf.level=0.95)
```

quantile_confint_nyblom

Two-sided quantile confidence interval based on interpolating the order statistic as suggested in Nyblom (1991)

Description

Two-sided quantile confidence interval based on interpolating the order statistic as suggested in Nyblom (1991)

Usage

```
quantile_confint_nyblom(
  x,
  p,
  conf.level = 0.95,
  x_is_sorted = FALSE,
  interpolate = TRUE,
  fix_interval = TRUE
)
```

Arguments

x	vector of observations
p	quantile of interest, $0 \leq p \leq 1$
conf.level	A conf.level * 100% confidence interval is computed
x_is_sorted	Boolean (Default: FALSE) indicating if x is already sorted and, hence, it is not necessary to sort it again. This is merely for speed reasons in situations where it is more efficient to only sort x once.
interpolate	Boolean (Default: TRUE) stating whether to interpolate the order statistics. If no interpolation is selected then this is just the standard exact procedure based on the order statistics. Note: This procedure is conservative (i.e. coverage is usually larger than the nominal conf.level and hence the interval is actually in general too large).
fix_interval	Boolean (Default: TRUE) For the case with no interpolation, try to extend interval upwards if coverage is too little.

Details

The interpolation procedure suggested by Nyblom (1992), which extends work by Hettmansperger and Sheather (1986) for the median is applied to the order statistic.

Value

A vector of length two containing the lower and upper limit of the confidence interval

References

Nyblom J, Note in interpolated order statistics, *Statistics and Probability Letters* 14, p. 129-131.

Examples

```
set.seed(123)
x <- rnorm(25)
quantile_confint_nyblom(x=x, p=0.8, conf.level=0.95, interpolate=TRUE)
```


Index

* **data**

flint, 2

flint, 2

median_confint_hs, 2

qci_coverage_one_sim, 4

quantile, 5

quantile_confint_boot, 5

quantile_confint_exact, 6

quantile_confint_nyblom, 7