Package 'statBasics'

March 16, 2022

Title Basic Functions to Statistical Methods Course

Version 0.1.0

Maintainer Gilberto Sassi <sassi.pereira.gilberto@gmail.com>

Description Basic statistical methods with some modifications for the course Statistical Methods at Federal University of Bahia (Brazil). All methods in this packages are explained in the text book of Montgomery and Runger (2010) <ISBN: 978-1-119-74635-5>.

Imports tibble, stats, stringr

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

Suggests testthat (>= 3.0.0), EnvStats, BSDA, purrr

Config/testthat/edition 3

NeedsCompilation no

Author Gilberto Sassi [aut, cre]

Repository CRAN

Date/Publication 2022-03-16 12:50:05 UTC

R topics documented:

ci_bern *Confidence Interval for Proportion*

Description

Confidence Interval for Proportion

Usage

```
ci_bern(x, n = NULL, conf_level = 0.95, type = "two.sided", na.rm = F)
```

Arguments

x	a vector of counts of successes.
n	a vector of counts of trials.
conf_level	confidence level of the returned confidence interval. Must be a single number between 0 and 1.
type	a character string specifying the type of confidence interval. Must be one of "two.sided" (default), "right" or "left".
na.rm	a logical value indicating whether NA values should be stripped before the computation proceeds.

Details

type specifies the type of confidence interval. If type is "two.sided", the returned confidence interval is (lower_ci,upper_ci). If type is "left", the returned confidence interval is (lower_ci,1). And, finally, is type is "right", the returned confidence interval is (0,upper_ci)).

Value

A 1 x 3 tibble with 'lower_ci', 'upper_ci' and 'conf_level' columns. Values correspond to lower, upper bounds of the confidence interval and confidence level, respectively.

Examples

```
heads <- rbinom(1, size = 100, prob = .5)
ci_bern(heads)
```

ci_exp *Confidence Interval of Exponential Distribution*

Description

Confidence Interval of Exponential Distribution

Usage

```
ci_exp(x, conf_level = 0.95, type = "two.sided", na.rm = F)
```

Arguments

x	a (non-empty) numeric vector.
conf_level	confidence level of the returned confidence interval. Must be a single number between 0 and 1.
type	a character string specifying the type of confidence interval. Must be one of "two.sided" (default), "right" or "less".
na.rm	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

"lower_ci" and "upper_ci" are computed using pivotal quantity, as explained by Montgomery and Runger «ISBN: 978-1-119-74635-5».

Value

A 1 x 3 tibble with 'lower_ci', 'upper_ci' and 'conf_level' columns. Values correspond to lower, upper bounds of the confidence interval and confidence level, respectively.

Examples

```
x <- rexp(1000)
ci_exp(x)
```

`ci_general` *Confidence Interval for Mean*

Description

Confidence Interval for Mean

Usage

```
ci_general(x, conf_level = 0.95, type = "two.sided", na.rm = F)
```

Arguments

`x`	a (non-empty) numeric vector.
`conf_level`	confidence level of the returned confidence interval. Must be a single number between 0 and 1.
`type`	a character string specifying the type of confidence interval. Must be one of "two.sided" (default), "right" or "less".
`na.rm`	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

"lower_ci" and "upper_ci" are computed using `t.test` function.

Value

A 1 x 3 tibble with 'lower_ci', 'upper_ci' and 'conf_level' columns. Values correspond to lower, upper bounds of the confidence interval and confidence level, respectively.

Examples

```
x <- rpois(1000, lambda = 10)
ci_general(x)
```

ci_norm *Confidence Interval for Normal Distribution*

Description

Confidence Interval for Normal Distribution

Usage

```
ci_norm(
  x,
  sd_pop = NULL,
  parameter = "mean",
  conf_level = 0.95,
  type = "two.sided",
  na.rm = F
)
```

Arguments

x	a (non-empty) numeric vector.
sd_pop	a number specifying the known standard deviation of the population.
parameter	a character string specifying the parameter in the normal distribution. Must be one of "mean" or "variance".
conf_level	confidence level of the returned confidence interval. Must be a single number between 0 and 1.
type	a character string specifying the type of confidence interval. Must be one of "two.sided" (default), "right" or "less".
na.rm	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

type specifies the type of confidence interval. If type is "two.sided", the returned confidence interval is (lower_ci,upper_ci) when parameter is "mean" or "variance". If type is "left", the returned confidence interval is (lower_ci,Inf) when parameter is "mean" or "variance". And, finally, is type is "right", the returned confidence interval is (-Inf,upper_ci)) when parameter is "mean", and the returned confidence interval is (0,upper_ci) when parameter is "variance".

Value

A 1 x 3 tibble with 'lower_ci', 'upper_ci' and 'conf_level' columns. Values correspond to lower, upper bounds of the confidence interval and confidence level, respectively.

Examples

```
x <- rnorm(1000)
ci_norm(x) # unknown variance and confidence interval for mean

x <- rnorm(1000, sd = 2)
ci_norm(x, sd_pop = 2) # known variance and confidence interval for mean

x <- rnorm(1000, sd = 5)
ci_norm(x, parameter = "variance") # confidence interval for variance
```

---

ht_1pop_mean *Hypothesis Mean for Normal Distribution*

Description

Hypothesis Mean for Normal Distribution

Usage

```
ht_1pop_mean(
  x,
  mu = 0,
  sd_pop = NULL,
  alternative = "two.sided",
  conf_level = NULL,
  sig_level = 0.05,
  na.rm = TRUE
)
```

Arguments

x	a (non-empty) numeric vector.
mu	a number indicating the true value of the mean. Default value is 0.
sd_pop	a number specifying the known standard deviation of the population. If sd_pop == NULL, we use the t-ttest. If !is.null(sd_pop), we use the z-test. Default value is NULL.
alternative	a character string specifying the alternative hypothesis, must be one of '"two.sided"' (default), '"greater"' or '"less"'. You can specify just the initial letter.
conf_level	a number indicating the confidence level to compute the confidence interval. If conf_level = NULL, then confidence interval is not included in the output. Default value is NULL.
sig_level	a number indicating the significance level to use in the General Procedure for Hypothesis Testing.
na.rm	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

I have wrapped the `t.test` and the `BSDA::z.test` in a function as explained in the book of Montgomery and Runger (2010) <ISBN: 978-1-119-74635-5>.

Value

a `tibble` with the following columns:

statistic the value of statistic.

p_value the p-value for the test.

critical_value critical value in the General Procedure for Hypothesis Testing.

critical_region critical region in the General Procedure for Hypothesis Testing.

mu a number indicating the true value of the mean.

alternative character string giving the direction of the alternative hypothesis.

lower_ci lower bound of the confidence interval. Is is present only if `!is.null(con_level)`.

upper_ci upper bound of the confidence interval. Is is present only if `!is.null(con_level)`.

Examples

```
sample <- rnorm(1000, mean = 10, sd = 2) #t-test
ht_1pop_mean(sample, mu = -1) # H0: mu == -1

sample <- rnorm(1000, mean = 5, sd = 3) # z-test
ht_1pop_mean(sample, mu = 0, sd_pop = 3, alternative = 'less') # H0: mu >= 0
```

---

ht_1pop_prop *Hypothesis Testing for Proportion*

Description

One-sample test to proportion.

Usage

```
ht_1pop_prop(
  x,
  n = NULL,
  proportion = 0.5,
  alternative = "two.sided",
  conf_level = NULL,
  sig_level = 0.05,
  na.rm = TRUE
)
```

Arguments

x	a (non-empty) numeric vector indicating the number of success. It can also be a vector with number of success, and it can be vector of 0 and 1.
n	a (non-empty) numeric vector indicating the number of trials. It can also be a vector with number of trials (if x is vector of success), and it can be NULL (if x is a vector of 0 e 1).
proportion	a number between 0 e 1 indicating the value in the null hypothesis. Default value is 0.5.
alternative	a character string specifying the alternative hypothesis, must be one of '"two.sided"' (default), '"greater"' or '"less"'. You can specify just the initial letter.
conf_level	a number indicating the confidence level to compute the confidence interval. If conf_level = NULL, then confidence interval is not included in the output. Default value is NULL.
sig_level	a number indicating the significance level to use in the General Procedure for Hypotheiss Testing.
na.rm	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

I have wrapped the book of Millard and Neerchal (2001) <ISBN: 978-0-367-39814-9>.

Value

a tibble with the following columns:

statistic the value of statistic.

p_value the p-value for the test.

critical_value critical value in the General Procedure for Hypothesis Testing.

critical_region critical region in the General Procedure for Hypothesis Testing.

proportion a number indicating the true value of the sigma.

alternative character string giving the direction of the alternative hypothesis.

lower_ci lower bound of the confidence interval. Is is present only if !is.null(con_level).

upper_ci upper bound of the confidence interval. Is is present only if !is.null(con_level).

Examples

```
amostra <- rbinom(1, size = 100, prob = 0.75)
ht_1pop_prop(amostra, proportion = 0.75, 100, conf_level = 0.99)

amostra <- c(rbinom(1, size = 10, prob = 0.75),
rbinom(1, size = 20, prob = 0.75),
rbinom(1, size = 30, prob = 0.75))
ht_1pop_prop(amostra, c(10, 20, 30), proportion = 0.99, conf_level = 0.90, alternative = 'less')

amostra <- rbinom(100, 1, prob = 0.75)
ht_1pop_prop(amostra, proportion = 0.01, conf_level = 0.95, alternative = 'greater')
```

ht_1pop_var *Hypothesis Testing for Variance*

Description

One-Sample chi-squared test on variance.

Usage

```
ht_1pop_var(
  x,
  sigma = 1,
  alternative = "two.sided",
  conf_level = NULL,
  sig_level = 0.05,
  na.rm = TRUE
)
```

Arguments

x	a (non-empty) numeric vector.
sigma	a number indicating the true value of the standard deviation in the null hypothesis. Default value is 1.
alternative	a character string specifying the alternative hypothesis, must be one of '"two.sided"' (default), '"greater"' or '"less"'. You can specify just the initial letter.
conf_level	a number indicating the confidence level to compute the confidence interval. If conf_level = NULL, then confidence interval is not included in the output. Default value is NULL.
sig_level	a number indicating the significance level to use in the General Procedure for Hypotheiss Testing.
na.rm	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

I have wrapped the EnvStats::varTest in a function as explained in the book of Montgomery and Runger (2010) <ISBN: 978-1-119-74635-5> and in the book of Millard and Neerchal (2001) <ISBN: 978-0-367-39814-9>.

Value

a tibble with the following columns:

statistic the value of statistic.

p_value the p-value for the test.

critical_value critical value in the General Procedure for Hypothesis Testing.

critical_region critical region in the General Procedure for Hypothesis Testing.

sigma a number indicating the true value of the sigma.

alternative character string giving the direction of the alternative hypothesis.

lower_ci lower bound of the confidence interval. Is is present only if `!is.null(con_level)`.

upper_ci upper bound of the confidence interval. Is is present only if `!is.null(con_level)`.

Examples

```
sample <- rnorm(1000, mean = 10, sd = 2)
ht_1pop_var(sample, sigma = 1) # H0: sigma = 1
```

Index