Package 'sprtt'

July 6, 2023

Type Package

Title Sequential Probability Ratio Tests Toolbox

Version 0.2.0

Maintainer Meike Steinhilber <Meike.Steinhilber@aol.com>

Description It is a toolbox for Sequential Probability Ratio Tests (SPRT), Wald (1945) <doi:10.2134/agronj1947.00021962003900070011x>. SPRTs are applied to the data during the sampling process, ideally after each observation. At any stage, the test will return a decision to either continue sampling or terminate and accept one of the specified hypotheses.
The seq_ttest() function performs one-sample, two-sample, and paired t-tests for testing one- and two-sided hypotheses (Schnuerch & Erdfelder (2019) <doi:10.1037/met0000234>).
The seq_anova() function allows to perform a sequential one-way fixed effects ANOVA (Steinhilber et al. (2023) <doi:10.31234/osf.io/m64ne>).
Learn more about the package by using vignettes ``browseVignettes(package = "sprtt``)" or go to the website <https://meikesteinhilber.github.io/sprtt/>.

License AGPL (>= 3)

URL https://meikesteinhilber.github.io/sprtt/

BugReports https://github.com/MeikeSteinhilber/sprtt/issues

Depends R (>= 3.5.0)

Imports methods, stats, dplyr, MBESS, purrr, glue, ggplot2, lifecycle

Suggests knitr, rmarkdown, testthat (>= 3.0.0), testthis, effsize, effectsize, vdiffr

VignetteBuilder knitr

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Author Meike Steinhilber [aut, cre] (<https://orcid.org/0000-0002-7144-2100>),
Martin Schnuerch [aut, ths] (<https://orcid.org/0000-0001-6531-2265>),
Anna-Lena Schubert [aut, ths] (<https://orcid.org/0000-0001-7248-0662>)

Repository CRAN

Date/Publication 2023-07-06 13:50:02 UTC

R topics documented:

`df_cancer` *Test data to run the examples*

Description

A dataset that includes 120 individuals.

Usage

```
df_cancer
```

Format

A data frame with 2 variables:

treatment_group

control_group

df_income *Test data to run the examples*

Description

A dataset that includes 120 individuals with sex gender and monthly income.

Usage

```
df_income
```

Format

A data frame with 2 variables:

monthly_income

sex

df_stress *Test data to run the examples*

Description

A dataset that includes 120 individuals.

Usage

```
df_stress
```

Format

A data frame with 2 variables:

baseline_stress

one_year_stress

`draw_sample_mixture` *Draw Samples from a Gaussian Mixture Distribution*

Description

[Experimental]

Draws exemplary samples with a certain effect size for the sequential one-oway ANOVA or the sequential t-test, see Steinhilber et al. (2023) doi:10.31234/osf.io/m64ne

Usage

```
draw_sample_mixture(k_groups, f, max_n, counter_n = 100, verbose = FALSE)
```

Arguments

`k_groups`	number of groups (levels of factor_A)
`f`	Cohen's f. The simulated effect size.
`max_n`	sample size for the groups (total sample size = max_n*k_groups)
`counter_n`	number of times the function tries to find a possible parameter combination for the distribution. Default value is set to 100.
`verbose`	`TRUE` or `FALSE`. Print out more information about the internal process of sampling the parameters (the internal counter that was reached, some additional hints and the drawn parameters for the Gaussian Mixture distributions.)

Value

returns a data.frame with the columns y (observations) and x (factor_A).

Examples

```
set.seed(333)

data <- sprtt::draw_sample_mixture(
  k_groups = 2,
  f = 0.40,
  max_n = 2
)
data

data <- sprtt::draw_sample_mixture(
  k_groups = 4,
  f = 1.2, # very large effect size
  max_n = 4,
  counter_n = 1000, # increase of counter is necessary
  verbose = TRUE # prints more information to the console
)
data
```

draw_sample_normal *Draw Samples from a Normal Distribution*

Description

[Experimental]

Draws exemplary samples with a certain effect size for the sequential one-oway ANOVA or the sequential t-test, see Steinhilber et al. (2023) doi:10.31234/osf.io/m64ne

Usage

```
draw_sample_normal(k_groups, f, max_n, sd = NULL, sample_ratio = NULL)
```

Arguments

k_groups	number of groups (levels of factor_A)
f	Cohen's f. The simulated effect size.
max_n	sample size for the groups (total sample size = max_n*k_groups)
sd	vector of standard deviations of the groups. Default value is 1 for each group.
sample_ratio	vector of sample ratios between th groups. Default value is 1 for each group.

Value

returns a data.frame with the columns y (observations) and x (factor_A).

Examples

```
set.seed(333)

data <- sprtt::draw_sample_normal(
  k_groups = 2,
  f = 0.20,
  max_n = 2
)
data

data <- sprtt::draw_sample_normal(
  k_groups = 4,
  f = 0,
  max_n = 2,
  sd = c(1, 2, 1, 8)
)
data

data <- sprtt::draw_sample_normal(
  k_groups = 3,
  f = 0.40,
```

```
  max_n = 2,
  sd = c(1, 0.8, 1),
  sample_ratio = c(1, 2, 3)
)
data
```

---

plot_anova *Plot Sequential ANOVA Results*

Description

[Experimental]

Creates plots for the results of the seq_anova() function.

Usage

```
plot_anova(
  anova_results,
  labels = TRUE,
  position_labels_x = 0.15,
  position_labels_y = 0.075,
  position_lr_x = 0.05,
  font_size = 25,
  line_size = 1.5,
  highlight_color = "#CD2626"
)
```

Arguments

anova_results	result object of the seq_anova() function (argument must be of class seq_anova_results).
labels	show labels in the plot.
position_labels_x	position of the boundary labels on the x-axis.
position_labels_y	position of the boundary labels on the y-axis.
position_lr_x	scales the position of the LR label on the x-axis.
font_size	font size of the plot.
line_size	line size of the plot.
highlight_color	highlighting color, default is "#CD2626" (red).

Value

returns a plot

Examples

```
# simulate data for the example ------------------------------------------------
set.seed(333)
data <- sprtt::draw_sample_normal(3, f = 0.25, max_n = 30)

# calculate the SPRT -----------------------------------------------------------
anova_results <- sprtt::seq_anova(y~x, f = 0.25, data = data, plot = TRUE)

# plot the results -------------------------------------------------------------
sprtt::plot_anova(anova_results)

sprtt::plot_anova(anova_results,
                  labels = TRUE,
                  position_labels_x = 0.5,
                  position_labels_y = 0.1,
                  position_lr_x = -0.5,
                  font_size = 25,
                  line_size = 2,
                  highlight_color = "green"
                  )

sprtt::plot_anova(anova_results,
                  labels = FALSE
                  )
```

---

seq_anova *Sequential Analysis of Variance*

Description

[Experimental]

Performs a sequential one-way fixed effects ANOVA, see Steinhilber et al. (2023) doi:10.31234/osf.io/m64ne for more information.

Usage

```
seq_anova(
  formula,
  f,
  alpha = 0.05,
  power = 0.95,
  data,
  verbose = TRUE,
  plot = FALSE,
  seq_steps = "single"
)
```

Arguments

formula	A formula specifying the model.
f	Cohen's f (expected minimal effect size or effect size of interest).
alpha	the type I error. A number between 0 and 1.
power	1 - beta (beta is the type II error probability). A number between 0 and 1.
data	A data frame in which the variables specified in the formula will be found.
verbose	a logical value whether you want a verbose output or not.
plot	calculates the ANOVA sequentially on the data and saves the results in the slot called plot. This calculation is necessary for the plot_anova() function.
seq_steps	Defines the sequential steps for the sequential calculation if `plot = TRUE`. Argument takes either a vector of numbers or the argument `single` or `balanced`. A vector of numbers specifies the sample sizes at which the anova is calculated. `single` specifies that after each single point the test statistic is calculated (step size = 1). Attention: the calculation starts at the number of groups times two. If the data do not fit to this, you have to specify the sequential steps yourself in this argument. `balanced` specifies that the step size is equal to the number of groups. Attention: the calculation starts at the number of groups times two.

Value

An object of the S4 class `seq_anova_results`. Click on the class link to see the full description of the slots. To get access to the object use the `@`-operator or `[]`-brackets instead of `$`. See the examples below.

Examples

```
# simulate data ------------------------------------------------------------
set.seed(333)
data <- sprtt::draw_sample_normal(k_groups = 3,
                    f = 0.25,
                    sd = c(1, 1, 1),
                    max_n = 50)


# calculate sequential ANOVA -----------------------------------------------
results <- sprtt::seq_anova(y ~ x, f = 0.25, data = data)
# test decision
results@decision
# test results
results

# calculate sequential ANOVA -----------------------------------------------
results <- sprtt::seq_anova(y ~ x,
                            f = 0.25,
                            data = data,
                            alpha = 0.01,
                            power = .80,
                            verbose = TRUE)
```

```
results

# calculate sequential ANOVA ----------------------------------------------------
results <- sprtt::seq_anova(y ~ x,
                            f = 0.15,
                            data = data,
                            alpha = 0.05,
                            power = .80,
                            verbose = FALSE)
results
```

`seq_ttest` *Sequential Probability Ratio Test using t-statistic*

Description

Performs one and two sample sequential t-tests on vectors of data. For more information on the sequential t-test, see Schnuerch & Erdfelder (2019) doi:10.1037/met0000234.

Usage

```
seq_ttest(
  x,
  y = NULL,
  data = NULL,
  mu = 0,
  d,
  alpha = 0.05,
  power = 0.95,
  alternative = "two.sided",
  paired = FALSE,
  na.rm = TRUE,
  verbose = TRUE
)
```

Arguments

`x`	Works with two classes: `numeric` and `formula`. Therefore you can write `"x"` or `"x~y"`. • `"numeric input"`: a (non-empty) numeric vector of data values. • `"formula input"`: a formula of the form lhs ~ rhs where lhs is a numeric variable giving the data values and rhs either 1 for a one-sample test or a factor with two levels giving the corresponding groups.
`y`	an optional (non-empty) numeric vector of data values.
`data`	an optional `data.frame`, which you can use only in combination with a `"formula input"` in argument `x`.

mu	a number indicating the true value of the mean (or difference in means if you are performing a two sample test).
d	a number indicating the specified effect size (Cohen's d)
alpha	the type I error. A number between 0 and 1.
power	1 - beta (beta is the type II error probability). A number between 0 and 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.
paired	a logical indicating whether you want a paired t-test.
na.rm	a logical value indicating whether `NA` values should be stripped before the computation proceeds.
verbose	a logical value whether you want a verbose output or not.

Value

An object of the S4 class `seq_ttest_results`. Click on the class link to see the full description of the slots. To get access to the object use the `@`-operator or `[]`-brackets instead of `$`. See the examples below.

Examples

```
# set seed ------------------------------------------------------------------
set.seed(333)

# load library --------------------------------------------------------------
library(sprtt)

# one sample: numeric input -------------------------------------------------
treatment_group <- rnorm(20, mean = 0, sd = 1)
results <- seq_ttest(treatment_group, mu = 1, d = 0.8)

# get access to the slots ---------------------------------------------------
# @ Operator
results@likelihood_ratio

# [] Operator
results["likelihood_ratio"]

# two sample: numeric input-------------------------------------------------
treatment_group <- stats::rnorm(20, mean = 0, sd = 1)
control_group <- stats::rnorm(20, mean = 1, sd = 1)
seq_ttest(treatment_group, control_group, d = 0.8)

# two sample: formula input -------------------------------------------------
stress_level <- stats::rnorm(20, mean = 0, sd = 1)
sex <- as.factor(c(rep(1, 10), rep(2, 10)))
seq_ttest(stress_level ~ sex, d = 0.8)

# NA in the data ------------------------------------------------------------
stress_level <- c(NA, stats::rnorm(20, mean = 0, sd = 2), NA)
```

```
sex <- as.factor(c(rep(1, 11), rep(2, 11)))
seq_ttest(stress_level ~ sex, d = 0.8, na.rm = TRUE)

# work with dataset (data are in the package included) ------------------------
seq_ttest(monthly_income ~ sex, data = df_income, d = 0.8)
```

Index