

Package ‘dabestr’

October 13, 2023

Type Package

Title Data Analysis using Bootstrap-Coupled Estimation

Version 2023.9.12

Description Data Analysis using Bootstrap-Coupled ESTimation.

Estimation statistics is a simple framework that avoids the pitfalls of significance testing. It uses familiar statistical concepts: means, mean differences, and error bars. More importantly, it focuses on the effect size of one's experiment/intervention, as opposed to a false dichotomy engendered by P values.

An estimation plot has two key features:

1. It presents all datapoints as a swarmplot, which orders each point to display the underlying distribution.
2. It presents the effect size as a bootstrap 95% confidence interval on a separate but aligned axes.

Estimation plots are introduced in Ho et al., Nature Methods 2019, 1548-7105.

<[doi:10.1038/s41592-019-0470-3](https://doi.org/10.1038/s41592-019-0470-3)>.

The free-to-view PDF is located at <https://www.nature.com/articles/s41592-019-0470-3.epdf?author_access_token=Euy6APITxsYA3huBKOFBvNRgN0jAjWe19jnR3ZoTv0Pr6zJiJ3AA5aH4989g0JS_dajtNr1Wt17D0fh-t4GFcvqwMYN03qb8C33na_UrCUCgrt-Z0J9aPL6TPSb0xIC-pbHWKUDo2XsUOr3hQm1Rew%3D%3D>.

<https://www.nature.com/articles/s41592-019-0470-3.epdf?author_access_token=Euy6APITxsYA3huBKOFBvNRgN0jAjWe19jnR3ZoTv0Pr6zJiJ3AA5aH4989g0JS_dajtNr1Wt17D0fh-t4GFcvqwMYN03qb8C33na_UrCUCgrt-Z0J9aPL6TPSb0xIC-pbHWKUDo2XsUOr3hQm1Rew%3D%3D>.

License Apache License (>= 2)

Encoding UTF-8

URL <https://github.com/ACCLAB/dabestr>,

<https://acclab.github.io/dabestr/>

Depends R (>= 2.10)

Imports ggplot2, cowplot, tidyr, dplyr, tibble, rlang, magrittr,
ggbeeswarm, effsize, grid, scales, ggsci, cli, boot, stats,
stringr, brunnermunzel, methods

RoxygenNote 7.2.3

Suggests testthat (>= 3.0.0), vdiffr, knitr, rmarkdown, kableExtra

Config/testthat/edition 3

LazyData true**VignetteBuilder** knitr, kableExtra**NeedsCompilation** no

Author Joses W. Ho [aut] (<<https://orcid.org/0000-0002-9186-6322>>),
 Kah Seng Lian [aut],
 Zhuoyu Wang [aut],
 Jun Yang Liao [aut],
 Felicia Low [aut],
 Tayfun Tumkaya [aut] (<<https://orcid.org/0000-0001-8425-3360>>),
 Yishan Mai [cre, ctb] (<<https://orcid.org/0000-0002-7199-380X>>),
 Sangyu Xu [ctb] (<<https://orcid.org/0000-0002-4927-9204>>),
 Hyungwon Choi [ctb] (<<https://orcid.org/0000-0002-6687-3088>>),
 Adam Claridge-Chang [ctb] (<<https://orcid.org/0000-0002-4583-3650>>),
 ACCLAB [cph, fnd]

Maintainer Yishan Mai <mai.yishan@u.duke.nus.edu>**Repository** CRAN**Date/Publication** 2023-10-13 11:50:06 UTC

R topics documented:

dabest_plot	2
effect_size	3
load	5
plot_kwargs	7
Index	9

dabest_plot	<i>Producing an estimation plot</i>
-------------	-------------------------------------

Description

Produces a Gardner-Altman estimation plot or a Cumming estimation plot depending on whether `float_contrast` is TRUE. The plot presents all datapoints as a swarmplot, which orders each point to display the underlying distribution. It also presents the effect size as a bootstrap 95% confidence interval (95% CI) on a separate but aligned axes.

Usage

```
dabest_plot(dabest_effectsize_obj, float_contrast = TRUE, ...)
```

Arguments

- `dabest_effectsize_obj` A `dabest_effectsize_obj` created by loading in a `dabest_obj` along with other specified parameters with the `effect_size()` function.
- `float_contrast` Default TRUE. If TRUE, a Gardner-Altman plot will be produced. If FALSE, a Cumming estimation plot will be produced.
- ... Adjustment parameters to control and adjust the appearance of the plot. (list of all possible adjustment parameters can be found under `plot_kwargs`)

Examples

```
# Loading of the dataset
data(twogroup_data)

# Preparing the data to be plotted
dabest_obj <- load(non_proportional_data,
  x = Group, y = Measurement,
  idx = c("Control 1", "Test 1")
)
dabest_obj.mean_diff <- mean_diff(dabest_obj)

# Plotting an estimation plot
dabest_plot(dabest_obj.mean_diff, TRUE)
```

effect_size

Calculating effect sizes

Description

Computes the effect size for each control-test group pairing in `idx`. The resampling bootstrap distribution of the effect size is then subjected to Bias-corrected and accelerated bootstrap (BCa) correction.

The following effect sizes `mean_diff`, `median_diff`, `cohens_d`, `hedges_g` and `cliffs_delta` are used for most plot types.

Usage

```
mean_diff(dabest_obj, perm_count = 5000)

median_diff(dabest_obj, perm_count = 5000)

cohens_d(dabest_obj, perm_count = 5000)

hedges_g(dabest_obj, perm_count = 5000)

cliffs_delta(dabest_obj, perm_count = 5000)

cohens_h(dabest_obj, perm_count = 5000)
```

Arguments

dabest_obj	A dabest_obj created by loading in dataset along with other specified parameters with the <code>load()</code> function.
perm_count	The number of reshuffles of control and test labels to be performed for each p-value.

Details

The plot types listed under here are limited to use only the following effect sizes.

- Proportion plots offers only mean_diff and cohens_h.
- Mini-Meta Delta plots offers only mean_diff.

The other plots are able to use all given basic effect sizes as listed in the Description.

Value

Returns a dabest_effectsize_obj list with 22 elements. The following are the elements contained within:

- raw_data The tidy dataset passed to `load()` that was cleaned and altered for plotting.
- idx The list of control-test groupings as initially passed to `load()`.
- delta_x_labels Vector containing labels for the x-axis of the delta plot.
- delta_y_labels String label for the y-axis of the delta plot.
- Ns List of labels for x-axis of the raw plot.
- raw_y_labels Vector containing labels for the y-axis of the raw plot.
- is_paired Boolean value determining if it is a paired plot.
- is_colour Boolean value determining if there is a colour column for the plot.
- paired Paired ("sequential" or "baseline") as initially passed to `load()`.
- resamples The number of resamples to be used to generate the effect size bootstraps.
- control_summary Numeric value for plotting of control summary lines for float_contrast = TRUE.
- test_summary Numeric value for plotting of control summary lines for float_contrast = TRUE.
- ylim Vector containing the y limits for the raw plot.
- enquox Quosure of x as initially passed to `load()`.
- enquoy Quosure of y as initially passed to `load()`.
- enquoidcol Quosure of id_col as initially passed to `load()`.
- enquocolour Quosure of colour as initially passed to `load()`.
- proportional Boolean value as initially passed to `load()`.
- minimeta Boolean value as initially passed to `load()`.
- delta Boolean value as initially passed to `load()`.
- proportional_data List of calculations related to the plotting of proportion plots.

- `boot_result` List containing values related to the calculation of the effect sizes, bootstrapping and BCa correction.
- `baseline_ec_boot_result` List containing values related to the calculation of the effect sizes, bootstrapping and BCa correction for the baseline error curve.
- `permtest_pvals` List containing values related to the calculations of permutation t tests and the corresponding p values, and p values for different types of effect sizes and different statistical tests.

Examples

```
# Loading of the dataset
data(non_proportional_data)

# Applying effect size to the dabest object
dabest_obj <- load(non_proportional_data,
  x = Group, y = Measurement,
  idx = c("Control 1", "Test 1")
)
dabest_obj.mean_diff <- mean_diff(dabest_obj)

# Printing dabest effectsize object
print(dabest_obj.mean_diff)
```

load

Loading data with dabestr

Description

Processes and converts a tidy dataset into the dabestr format. The output of this function is then used as an input for various procedural functions within dabestr to create estimation plots.

Usage

```
load(
  data,
  x,
  y,
  idx = NULL,
  paired = NULL,
  id_col = NULL,
  ci = 95,
  resamples = 5000,
  colour = NULL,
  proportional = FALSE,
  minimeta = FALSE,
  delta2 = FALSE,
  experiment = NULL,
```

```

    experiment_label = NULL,
    x1_level = NULL
  )

```

Arguments

<code>data</code>	A tidy dataframe.
<code>x</code>	Column in data that contains the treatment groups.
<code>y</code>	Column in data that contains the measurement values.
<code>idx</code>	List of control-test groupings for which the effect size will be computed for.
<code>paired</code>	Paired ("sequential" or "baseline"). Used for plots for experiments with repeated-measures designs. If "sequential", comparison happens between each measurement to the one directly preceding it. (control vs group i) If "baseline", comparison happens between each group to a shared control. (group i vs group i+1)
<code>id_col</code>	Column in data indicating the identity of the datapoint if the data is tagged. Compulsory parameter if paired is TRUE.
<code>ci</code>	Default 95. Determines the range of the confidence interval for effect size and bootstrap calculations. Only accepts values between 0 to 100 (inclusive).
<code>resamples</code>	The number of resamples to be used to generate the effect size bootstraps.
<code>colour</code>	Column in data that determines the groupings for colour of the swarmplot as opposed to x.
<code>proportional</code>	Boolean value determining if proportion plots are being produced.
<code>minimeta</code>	Boolean value determining if mini-meta analysis is conducted.
<code>delta2</code>	Boolean value determining if delta-delta analysis for 2 by 2 experimental designs is conducted.
<code>experiment</code>	Experiment column name for delta-delta analysis.
<code>experiment_label</code>	String specifying the experiment label that is used to distinguish the experiment and the factors (being used in the plotting labels).
<code>x1_level</code>	String setting the first factor level in a 2 by 2 experimental design.

Value

Returns a `dabest_obj` list with 18 elements. The following are the elements contained within:

- `raw_data` The tidy dataset passed to `load()` that was cleaned and altered for plotting.
- `proportional_data` List of calculations related to the plotting of proportion plots.
- `enquo_x` Quosure of x as initially passed to `load()`.
- `enquo_y` Quosure of y as initially passed to `load()`.
- `enquo_id_col` Quosure of `id_col` as initially passed to `load()`.
- `enquo_colour` Quosure of `colour` as initially passed to `load()`.

- `proportional` Boolean value determining if proportion plots are being produced.
- `minimeta` Boolean value determining if mini-meta analysis is conducted.
- `delta2` Boolean value determining if delta-delta analysis for 2 by 2 experimental designs is conducted.
- `idx` List of control-test groupings for which the effect size will be computed for.
- `resamples` The number of resamples to be used to generate the effect size bootstraps.
- `is_paired` Boolean value determining if it is a paired plot.
- `is_colour` Boolean value determining if there is a specified colour column for the plot.
- `paired` Paired ("sequential" or "baseline") as initially passed to `load()`.
- `ci` Numeric value which determines the range of the confidence interval for effect size and bootstrap calculations. Only accepts values between 0 to 100 (inclusive).
- `Ns` List of labels for x-axis of the rawdata swarm plot.
- `control_summary` Numeric value for plotting of control summary lines for `float_contrast=TRUE`.
- `test_summary` Numeric value for plotting of test summary lines for `float_contrast = TRUE`.
- `ylim` Vector containing the y limits for the rawdata swarm plot.

Examples

```
# Loading in of the dataset
data(non_proportional_data)

# Creating a dabest object
dabest_obj <- load(
  data = non_proportional_data, x = Group, y = Measurement,
  idx = c("Control 1", "Test 1")
)

# Printing dabest object
print(dabest_obj)
```

plot_kwarg

Adjustable Plot Aesthetics

Description

These are the available plot kwarg for adjusting the plot aesthetics of your estimation plot:

- `swarm_label` Default "value" or "proportion of success" for proportion plots. Label for the y-axis of the swarm plot.
- `contrast_label` Default "effect size", based on the effect sizes as given in `effect_size()`. Label for the y-axis of the contrast plot.
- `delta2_label` Default NULL. Label for the y-label for the delta-delta plot.

- `swarm_x_text` Default 11. Numeric value determining the font size of the x-axis of the swarm plot.
- `swarm_y_text` Default 15. Numeric value determining the font size of the y-axis of the swarm plot.
- `contrast_x_text` Default 11. Numeric value determining the font size of the x-axis of the delta plot.
- `contrast_y_text` Default 15. Numeric value determining the font size of the y-axis of the delta plot.
- `swarm_ylim` Default NULL. Vector containing the y limits for the swarm plot
- `contrast_ylim` Default NULL. Vector containing the y limits for the delta plot.
- `delta2_ylim` Default NULL. Vector containing the y limits for the delta-delta plot.
- `raw_marker_size` Default 1.5. Numeric value determining the size of the points used in the swarm plot.
- `tufte_size` Default 0.8. Numeric value determining the size of the tufte line in the swarm plot.
- `es_marker_size` Default 0.5. Numeric value determining the size of the points used in the delta plot.
- `es_line_size` Default 0.8. Numeric value determining the size of the ci line in the delta plot.
- `raw_marker_alpha` Default 1. Numeric value determining the transparency of the points in the swarm plot.
- `raw_bar_width` Default 0.3. Numeric value determining the width of the bar in the sankey diagram.
- `raw_marker_spread` Default 2. The distance between the points if it is a swarm plot.
- `raw_marker_side_shift` Default 0. The horizontal distance that the swarm plot points are moved in the direction of the `asymmetric_side`.
- `asymmetric_side` Default "right". Can be either "right" or "left". Controls which side the swarm points are shown.
- `show_delta2` Default FALSE. Boolean value determining if the delta-delta plot is shown.
- `show_mini_meta` Default FALSE. Boolean value determining if the weighted average plot is shown. If False, the resulting graph would be identical to a multiple two-groups plot.
- `show_zero_dot` Default TRUE. Boolean value determining if there is a dot on the zero line of the effect size for the control-control group.
- `show_baseline_ec` Default FALSE. Boolean value determining whether the baseline curve is shown.
- `sankey` Default TRUE. Boolean value determining if the flows between the bar charts will be plotted.
- `raw_flow_alpha` Default 0.5. Numeric value determining the transparency of the sankey flows in a paired proportion plot.
- `flow` Default TRUE. Boolean value determining whether the bars will be plotted in pairs.
- `custom_palette` Default "d3". String. The following palettes are available for use: `npg`, `aaas`, `nejm`, `lancet`, `jama`, `jco`, `ucscgb`, `d3`, `locuszoom`, `igv`, `cosmic`, `uchicago`, `brewer`, `ordinal`, `viridis_d`.

Index

`cliffs_delta(effect_size)`, 3

`cohens_d(effect_size)`, 3

`cohens_h(effect_size)`, 3

`dabest_plot`, 2

`effect_size`, 3

`effect_size()`, 3, 7

`hedges_g(effect_size)`, 3

`load`, 5

`load()`, 4, 6, 7

`mean_diff(effect_size)`, 3

`median_diff(effect_size)`, 3

`plot_kwargs`, 3, 7