Package ‘dabestr’

February 26, 2025

Type Package

Title Data Analysis using Bootstrap-Coupled Estimation

Version 2025.3.14

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>. The free-to-view PDF is located at <`https://www.nature.com/articles/s41592-019-0470-3.epdf?author_access_token=Euy6APITxsYA3huBKOFBvNRgN0jAjWel9jnR3ZoTv0Pr6zJiJ3AA5aH4989gOJS_dajtNr1Wt17D0fh-t4GFcvqwMYN03qb8C33na_UrCUcGrt-Z0J9aPL6TPSbOxIC-pbHWKUDo2XsUOr3hQmlRew%3D%3D`>.

License Apache License (>= 2)

URL `https://github.com/ACCLAB/dabestr`, `https://acclab.github.io/dabestr/`

Depends R (>= 2.10)

Imports boot, brunnermunzel, cli, cowplot, dplyr, effsize, ggbeeswarm, ggplot2 (>= 3.5.1), ggsci, grid, magrittr, RColorBrewer, rlang, scales, stats, stringr, tibble, tidyr, viridisLite

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

VignetteBuilder kableExtra, knitr

Config/testthat/edition 3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Author Joses W. Ho [aut] (<https://orcid.org/0000-0002-9186-6322>),
Kah Seng Lian [aut],
Ana Rosa Castillo [aut],
Zhuoyu Wang [aut],
Jun Yang Liao [aut],
Felicia Low [aut],
Tayfun Tumkaya [aut] (<https://orcid.org/0000-0001-8425-3360>),
Jonathan Anns [ctb] (<https://orcid.org/0009-0005-8349-4986>),
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 <`maiyishan@u.duke.nus.edu`>

License_is_FOSS yes

Repository CRAN

Date/Publication 2025-02-26 12:50:02 UTC

Contents

`dabest_plot` *Producing an estimation plot*

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 `load()` 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.
- `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 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)
```

`forest_plot` *Generates a Forest Plot*

Description

This function creates a forest plot summarizing a list of contrasts.

Usage

```
forest_plot(
  contrasts,
  contrast_labels,
  contrast_type = "delta2",
  effect_size = "mean_diff",
  ylabel = "effect size",
  title = "Delta Delta Forest",
  fontsize = 12,
  title_font_size = 16,
  violin_kwargs = NULL,
  marker_size = 1.1,
  ci_line_width = 1.3,
  custom_palette = NULL,
  rotation_for_xlabels = 0,
  alpha_violin_plot = 0.8
)
```

Arguments

`contrasts`	A list of contrast objects. These objects should contain the statistical information for each comparison (e.g., estimates, standard errors).
`contrast_labels`	A list of labels for the contrast objects. E.g., c(’Drug1’, ’Drug2’, ’Drug3’) These labels will be used to identify each comparison on the plot.
`contrast_type`	Select between "delta2" (for delta-delta) or "minimeta" for mini-meta analysis. This determines the type of effect size calculation used in the plot.
`effect_size`	Character string specifying the effect size metric to display. Valid options include "mean_diff", "median_diff", "cliffs_delta", "cohens_d", "hedges_g", or "delta_g". The default is "mean_diff".
`ylabel`	Character string specifying the axis label for the dependent variable (Y-axis for vertical layout, X-axis for horizontal layout). The default is "value".
`title`	Character string specifying the title for the forest plot. The default is "Delta delta Forest".
`fontsize`	Font size for text elements in the plot. Default is 12.
`title_font_size`	Font size for text of plot title. Defaults is 16.
`violin_kwargs`	Additional arguments for violin plot customization. Default is NULL
`marker_size`	Marker size for plotting mean differences or effect sizes. Default is 20.
`ci_line_width`	Width of confidence interval lines. Default is 2.5.
`custom_palette`	A list or key:value pair of colors, one for each contrast object. E.g., c(’gray’, ’blue’, ’green’) or c(’Drug1’=’gray’, ’Drug2’=’blue’, ’Drug3’=’green’). Default NULL.
`rotation_for_xlabels`	Rotation angle for x-axis labels, improving readability. Default is 45.
`alpha_violin_plot`	Transparency level for violin plots. Default is 0.8

Value

A ggplot object representing the forest plot.

`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

`data`	A tidy dataframe.
`x`	Column in `data` that contains the treatment groups.
`y`	Column in `data` that contains the measurement values.
`idx`	List of control-test groupings for which the effect size will be computed for.
`paired`	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)
`id_col`	Column in `data` indicating the identity of the datapoint if the data is tagged. Compulsory parameter if paired is TRUE.
`ci`	Default 95. Determines the range of the confidence interval for effect size and bootstrap calculations. Only accepts values between 0 to 100 (inclusive).
`resamples`	The number of resamples to be used to generate the effect size bootstraps.
`colour`	Column in `data` that determines the groupings for colour of the swarmplot as opposed to `x`.
`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.
`experiment`	Experiment column name for delta-delta analysis.

`experiment_label`	String specifying the experiment label that is used to distinguish the experiment and the factors (being used in the plotting labels).
`x1_level`	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_kwargs` *Adjustable Plot Aesthetics*

Description

These are the available plot kwargs 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.
- `show_legend` Default TRUE. If TRUE, legend will be shown. If FALSE, legend will not be 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.
- `contrast_bars` Default TRUE. Whether or not to display the contrast bars at the delta plot.
- `params_contrast_bars`. Default value: list(color = NULL, alpha = 0.3). Pass relevant keyword arguments to the contrast bars.
- `swarm_bars` Default TRUE. Whether or not to display the swarm bars.
- `params_swarm_bars`. Default value: list(color = NULL, alpha = 0.3). Pass relevant keyword arguments to the swarm bars.

Index