Package 'steppedwedge'

October 30, 2026

Type Package

Title Analyze Data from Stepped Wedge Cluster Randomized Trials

Version 1.1.0

Description Provide various functions and tools to help fit models for estimating treatment effects in stepped wedge cluster randomized trials. Implements methods described in Kenny, Voldal, Xia, and Heagerty (2022) ``Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect", <doi:10.1002/sim.9511>.

License GPL-3

Encoding UTF-8

LazyData true

VignetteBuilder knitr

Imports dplyr (>= 1.0.10), stats (>= 4.0.0), forcats (>= 1.0.0), stringr (>= 1.5.1), magrittr (>= 2.0.3), methods (>= 3.6.2), ggplot2 (>= 3.5.1), lme4 (>= 2.0.0), geepack (>= 1.3.12), performance (>= 0.12.4), Matrix (>= 1.7.1), MASS (>= 7.3.61), expm (>= 1.0.0)

URL https://avi-kenny.github.io/steppedwedge/, https://github.com/Avi-Kenny/steppedwedge

BugReports https://github.com/Avi-Kenny/steppedwedge/issues

Suggests knitr, rmarkdown, testthat (>= 3.0.0), tidyr

Config/testthat/edition 3

Depends R (>= 3.5)

Config/roxygen2/version 8.0.0

RoxygenNote 7.3.3

NeedsCompilation no

Author Avi Kenny [aut, cre, cph], David Arthur [aut], Yongdong Ouyang [aut]

Maintainer Avi Kenny <avi.kenny@gmail.com>

Repository CRAN

Date/Publication 2026-10-30 19:20:02 UTC

Contents

analyze 2
as_table 5
load_data 5
params 7
plot_clusters 8
plot_design 9
plot_effect_curves 10
sw_data_example 11
sw_data_example_binom 11

Index **13**

`analyze` *Analyze a stepped wedge dataset*

Description

Analyze a stepped wedge dataset

Usage

```
analyze(
  dat,
  method = "mixed",
  estimand_type = "TATE",
  estimand_time = c(1, max(dat$exposure_time)),
  exp_time = "IT",
  cal_time = "categorical",
  family = stats::gaussian,
  exponentiate = FALSE,
  re = c("clust", "time"),
  corstr = "exchangeable",
  advanced = params()
)
```

Arguments

`dat`	A dataframe containing the stepped wedge trial data.
`method`	A character string; either "mixed", for a mixed-effects model, or "GEE", for generalized estimating equations.
`estimand_type`	One of c("TATE", "PTE"); "TATE" represents the time-averaged treatment effect and "PTE" represents the point treatment effect.

`estimand_time` An integer vector of length 1 or 2. When estimand_type="TATE", 'estimand_time' must be a numeric vector of length 2, representing the start and end times of the exposure time period to average over. When estimand_type="PTE", 'estimand_time' must be a numeric vector of length 1, representing the time period of interest. See examples.

`exp_time` One of c("IT", "ETI", "DCT", "NCS", "TEH"); model for exposure time. "IT" encodes an immediate treatment model with a single treatment effect parameter. "ETI" is an exposure time indicator model, including one indicator variable for each exposure time point. "NCS" uses a natural cubic spline model for the exposure time trend. "TEH" includes a random slope term in the model, allowing the treatment effect to vary by timepoint. "DCT" encodes a delayed constant treatment model, which estimates individual effects for a specified number of washout periods (set via the 'w' parameter in 'advanced') followed by a single constant treatment effect.

`cal_time` One of c("categorical", "NCS", "linear", "none"); model for calendar time. "categorical" uses indicator variables for discrete time points, as in the Hussey and Hughes model. "NCS" uses a natural cubic spline, useful for datasets with continuous time. "linear" uses a single slope parameter. "none" assumes that there is no underlying calendar time trend.

`family` A family object; see documentation for 'glm'.

`exponentiate` Logical; if TRUE, return exponentiated treatment effect estimates and confidence intervals (including in the 'effect_curve' object). Defaults to FALSE.

`re` A character vector of random effects to include; only relevant if method="mixed" is used. Possible random effects include "clust" (random intercept for cluster), "time" (random intercept for each cluster-period, i.e. each combination of cluster and calendar time; calendar time is the 'time' variable passed to 'load_data()', not exposure time), "ind" (random intercept for individuals; appropriate when a cohort design is used), and "tx" (a cluster-level random slope on treatment; see Details).

`corstr` One of c("independence", "exchangeable", "ar1"); only relevant if method="GEE" is used. Defines the GEE working correlation structure; see the documentation for 'geepack::geeglm'.

`advanced` A list of options returned by `params`.

Details

Covariates specified via the 'covariates' argument of `load_data` are included in the mean model as additive main effects, alongside the calendar time and exposure time terms. Covariates should typically be measured before randomization.

With 're = "tx"', each cluster has a random slope on the treatment indicator. This results in an additive deviation from the main treatment effect coefficient. In models that allow for a time-varying treatment effect (ETI, NCS, DCT, and TEH), the cluster-specific deviation results in an additive scaling of the entire effect curve. Here, "additive" refers to the linear predictor scale; the effects relative to the outcome will be different if a nonlinear link function is used. For example, if a log link function is used, the deviations will have a multiplicative effect on the outcome.

Value

A list with the model object, model type as a string, estimand type as a string, numeric treatment effect estimate, numeric treatment effect standard error, treatment effect 95 p-value corresponding to the null hypothesis that the main treatment effect estimand equals zero, a list with treatment effect estimates (and standard errors and 95 passed to 'analyze()', and an indicator whether the effect estimates and CI are exponentiated.

Examples

```
# Load data
test_data <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_cont", data = sw_data_example)

# Analysis example 1: TATE estimand for exposure times 1 through 4
results_tate <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "ETI")

results_tate

# Analysis example 2: PTE estimand for exposure time 3
results_pte <- analyze(dat = test_data, method = "mixed", estimand_type = "PTE",
estimand_time = 3, exp_time = "ETI")

results_pte

# Analysis example 3: TATE estimand for exposure times 1 through 4, Natural Cubic Splines model
results_tate_ncs <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "NCS", advanced = params(n_knots_exp = 4))

results_tate_ncs

# Analysis example 4: TATE estimand for exposure times 1 through 4 with binomial outcome data
# Load data
test_data_bin <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = c("numerator", "denominator"), data = sw_data_example_binom)

results_pte_bin <- analyze(dat = test_data_bin, family = binomial, method = "mixed",
estimand_type = "TATE", estimand_time = c(1, 4), exp_time = "ETI")

results_pte_bin

# Analysis example 5: TATE estimand using the Delayed Constant Treatment (DCT) model,
# with a washout period of w = 2 exposure times
results_tate_dct <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "DCT", advanced = params(w = 2))

results_tate_dct
```

as_table *Create table of estimates*

Description

Format estimates returned by `analyze` as a table

Usage

```
as_table(..., labels = NA)
```

Arguments

`...`	One or more objects of class `"sw_analysis"` returned by `analyze`.
`labels`	A character vector of length equal to length(list(...)) representing curve labels

Value

A table of effect estimate values

Examples

```
# Load data
test_data <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_cont", data = sw_data_example)


IT_model <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "IT")
ETI_model <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "ETI")
NCS_model <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "NCS")
ests_table <- as_table(IT_model, IT_model, NCS_model)
head(ests_table)
```

load_data *Load and format data object*

Description

Load and format data object

Usage

```
load_data(
  time,
  cluster_id,
  individual_id = NULL,
  treatment,
  outcome,
  exposure_time = NULL,
  offset = NULL,
  time_type = "discrete",
  data,
  covariates = NULL
)
```

Arguments

time	A character string; the name of the numeric variable representing time. Time can be either discrete or continuous.
cluster_id	A character string; the name of the numeric variable identifying the cluster.
individual_id	A character string (optional); the name of the numeric variable identifying the individual.
treatment	A character string; the name of the binary variable indicating treatment. Values must be either integers (0/1) or Boolean (T/F).
outcome	Either a character string or a vector of two character strings; for a numeric or binary outcome, the single character string indicates the name of the numeric or binary outcome variable; for binomial outcome data, the vector of two character strings indicates the "# of successes" variable and the "# of trials" variable, respectively. Values in the outcome variable(s) must be either numeric or Boolean (T/F).
exposure_time	A character string (optional); the name of the numeric variable identifying the exposure time variable. If this is not provided, the package will calculate exposure time automatically based on `treatment` and `time`.
offset	A character string (optional); the name of the numeric variable specifying the offset. The offset is added to the linear predictor of every model fit by `analyze`. Rows with a missing offset are dropped.
time_type	One of c("discrete", "continuous"); whether the model treats time as discrete or continuous.
data	A dataframe containing the stepped wedge trial data.
covariates	A character vector (optional); the names of variables in ‘data‘ to include as additive covariates in the mean model fit by `analyze`. Covariates can be individual-level or cluster-level and must be numeric, binary, or categorical. Typically, covariates are chosen because of presumed association with the outcome, which may increase the precision of treatment effect estimation.

Value

An object of class `sw_data`

Examples

```
example_data <- load_data(time = "period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_cont", offset = NULL, data = sw_data_example)
base::summary(example_data)

example_data_binomial <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = c("numerator", "denominator"), data = sw_data_example_binom)
base::summary(example_data_binomial)

# Including covariates in the mean model
sw_data_cov <- sw_data_example
sw_data_cov$x <- stats::rnorm(nrow(sw_data_cov))
example_data_cov <- load_data(time = "period", cluster_id = "cluster",
treatment = "trt", outcome = "outcome_cont", data = sw_data_cov,
covariates = "x")
```

---

params *Set advanced parameters for analysis of data from stepped wedge trials*

Description

This should be used in conjunction with `analyze` to set parameters controlling analysis; see examples.

Usage

```
params(
  offset = NULL,
  n_knots_exp = 4,
  n_knots_cal = 4,
  var_est = "model",
  var_est_type = "classic",
  return_ncs = F,
  re_correlated = F,
  w = 1
)
```

Arguments

`offset`	Deprecated; use the ‘offset‘ argument of `load_data` instead.
`n_knots_exp`	An integer; only relevant when exp_time="NCS". Specifies the number of knots to use for exposure time, including boundary knots. The spline basis includes an intercept, and the degree of the basis is equal to the number of knots.
`n_knots_cal`	An integer; only relevant when cal_time="NCS". Specifies the number of knots to use for calendar time, including boundary knots. The spline basis includes an intercept, and the degree of the basis is equal to the number of knots.

var_est	A character string; either "model", for model-based variance, or "robust", to use the robust variance estimator.
var_est_type	A character string; one of c("classic","DF","KC","MD","FG"); only relevant when var_est="robust".
return_ncs	Logical; only relevant when exp_time="NCS". Specifies whether the full covariance matrix for the calendar time parameters and the transformed treatment effect parameters are returned.
re_correlated	Logical; specifies whether random treatment effect and random intercept for cluster are correlated.
w	Integer; the number of washout periods to use when ‘exp_time = "DCT"‘ (Delayed Constant Treatment). Defaults to 1.

Value

A list of options

Examples

```
dat <- load_data(time = "period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_bin", data = sw_data_example)

analyze(dat = dat, method = "mixed", estimand_type = "TATE",  exp_time = "NCS",
family = binomial, advanced = params(var_est = "robust"))
```

---

plot_clusters *Plot observed and predicted outcomes by cluster over time*

Description

Plot observed and predicted outcomes by cluster over time

Usage

```
plot_clusters(object, ncol = 3)
```

Arguments

object	An analysis object of class ‘sw_analysis‘ or a dataset of class ‘sw_data‘/‘sw_dat‘.
ncol	Integer; number of columns in the faceted plot. Defaults to 3.

Value

A list with a ‘ggplot2‘ object of the actual (and predicted, if applicable) outcomes by cluster.

Examples

```
# Load data
test_data <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_cont", data = sw_data_example)

# Plot just the dataset
plot_clusters(test_data)

# Analyze using TATE estimand for exposure times 1 through 4
results_tate <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "ETI")

# Plot analysis results (with predicted lines) by cluster
plot_clusters(results_tate)
```

---

plot_design *Plot stepped wedge design*

Description

Plot stepped wedge design

Usage

```
plot_design(dat)
```

Arguments

dat	A dataframe containing the stepped wedge trial data.

Value

A list with a plot of the stepped wedge design.

Examples

```
# Load data
example_data <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_cont", data = sw_data_example)
# Plot design
plot_design(example_data)
```

plot_effect_curves *Plot effect estimates by exposure time for one or more models.*

Description

Plot effect estimates by exposure time for one or more models.

Usage

```
plot_effect_curves(
  ...,
  labels = NA,
  facet_nrow = 1,
  continuous = FALSE,
  resolution = 0.1
)
```

Arguments

...	One or more objects of class "sw_analysis" returned by analyze.
labels	A character vector of length equal to the length of list(...), representing plot labels. Only used if length(list(...))>1.
facet_nrow	Number of rows for displaying plots using ggplot2::facet_wrap().
continuous	Logical; if TRUE, models fit with Natural Cubic Splines (exp_time = "NCS") will be evaluated and plotted as smooth, continuous curves. Models using discrete exposure times (e.g., ETI, DCT) will remain discrete.
resolution	Numeric; the step size for generating the continuous curve. Defaults to 0.1. Only relevant if continuous = TRUE.

Value

A plot of the effect curve for each "sw_analysis" object passed to the function.

Examples

```
# Load data
test_data <- load_data(time ="period", cluster_id = "cluster", individual_id = NULL,
treatment = "trt", outcome = "outcome_cont", data = sw_data_example)


IT_model <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "IT")
ETI_model <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "ETI")
NCS_4_model <- analyze(dat = test_data, method = "mixed", estimand_type = "TATE",
estimand_time = c(1, 4), exp_time = "NCS", advanced = params(n_knots_exp = 4))
```

```
# Plot discrete models
plot_effect_curves(IT_model, NCS_4_model, ETI_model, facet_nrow = 1)

# Plot continuous NCS model alongside discrete models
plot_effect_curves(IT_model, NCS_4_model, ETI_model, facet_nrow = 1, continuous = TRUE)
```

`sw_data_example` *Example stepped wedge data*

Description

Data generated for the purpose of demonstrating the steppedwedge package

Usage

```
sw_data_example
```

Format

'sw_data_example' A data frame with 2,063 rows and 5 columns:

cluster Cluster id

period Time period

trt Treatment indicator

outcome_bin Binary outcome

outcome_cont Continuous outcome ...

`sw_data_example_binom` *Example stepped wedge data with aggregrated outcomes*

Description

A simulated stepped wedge dataset formatted with separate columns for the number of successes (numerator) and number of trials (denominator).

Usage

```
sw_data_example_binom
```

Format

‘sw_data_example_binom‘ A data frame with 90 rows and 5 columns:

cluster Cluster id

period Time period

trt Treatment indicator

denominator Denominator (# trials)

numerator Numerator (# successes) ...

Index

∗ **datasets**
    sw_data_example, 11
    sw_data_example_binom, 11

analyze, 2, *5–7*, *10*
as_table, 5

load_data, *3*, 5, *7*

params, *3*, 7
plot_clusters, 8
plot_design, 9
plot_effect_curves, 10

sw_data_example, 11
sw_data_example_binom, 11