

Package ‘MRStdCRT’

October 8, 2026

Type Package

Title Model-Robust Standardization in Cluster-Randomized Trials

Version 0.1.2

Description Implements model-robust standardization for cluster-randomized trials (CRTs). Provides functions that standardize user-specified regression models to estimate marginal treatment effects. The targets include the cluster-average and individual-average treatment effects, with utilities for variance estimation and example simulation datasets. Methods are described in Li, Tong, Fang, Cheng, Kahan, and Wang (2025) <[doi:10.1002/sim.70270](https://doi.org/10.1002/sim.70270)>.

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 4.1)

Imports dplyr (>= 1.0.0), geepack (>= 1.3-2), lme4 (>= 1.1-25), nlme (>= 3.1-150), magrittr (>= 2.0.0), rlang (>= 1.0.0), stats

VignetteBuilder knitr

URL <https://github.com/deckardt98/MRStdCRT>

BugReports <https://github.com/deckardt98/MRStdCRT/issues>

Suggests knitr, rmarkdown

RoxygenNote 7.3.2

NeedsCompilation no

Author Jiaqi Tong [aut],
Changjun Li [aut, cre],
Xi Fang [aut],
Chao Cheng [aut],
Bingkai Wang [aut],
Fan Li [aut]

Maintainer Changjun Li <changjun.li@yale.edu>

Repository CRAN

Date/Publication 2026-10-08 02:20:02 UTC

Contents

data_sim_binary	2
data_sim_continuous	3
MRStdCRT_fit	4
MRStdCRT_point	6
ppact	7
summary.MRS_obj	8
Index	9

data_sim_binary	<i>Example Dataset: Simulated CRT (binary outcome)</i>
-----------------	--

Description

A simulated dataset for demonstrating **MRStdCRT** with a binary outcome. Treatment is assigned at the cluster level and is constant within cluster.

Usage

```
data(data_sim_binary)
```

Format

- A data frame with the following variables (10 columns):
- A** Cluster-level treatment assignment (0/1), constant within cluster.
 - H1** Cluster-level covariate 1.
 - H2** Cluster-level covariate 2.
 - N** Cluster size recorded on each row (repeats within cluster).
 - X1** Individual-level covariate 1 (numeric).
 - X2** Individual-level covariate 2 (numeric or binary coded 0/1).
 - Y** Observed binary outcome (0/1).
 - Y0** Potential outcome under control (0/1).
 - Y1** Potential outcome under treatment (0/1).
 - cluster_id** Cluster identifier (integer or factor), constant within cluster.

Source

Simulated data included with the package for examples.

Examples

```
data(data_sim_binary)
head(data_sim_binary)
with(data_sim_binary, table(A, Y))
```

data_sim_continuous	<i>Example Dataset: Simulated CRT (continuous outcome)</i>
---------------------	--

Description

A simulated dataset for demonstrating **MRStdCRT** with a continuous outcome. Treatment is assigned at the cluster level and is constant within cluster.

Usage

```
data(data_sim_continuous)
```

Format

A data frame with the following variables (10 columns):

A Cluster-level treatment assignment (0/1), constant within cluster.

H1 Cluster-level covariate 1.

H2 Cluster-level covariate 2.

N Cluster size recorded on each row (repeats within cluster).

X1 Individual-level covariate 1 (numeric).

X2 Individual-level covariate 2 (numeric or binary coded 0/1).

Y Observed continuous outcome.

Y0 Potential outcome under control (continuous).

Y1 Potential outcome under treatment (continuous).

cluster_id Cluster identifier (integer or factor), constant within cluster.

Source

Simulated data included with the package for examples.

Examples

```
data(data_sim_continuous)
head(data_sim_continuous)
table(data_sim_continuous$cluster_id)
```

MRStdCRT_fit

*Model-robust Standardization Estimators for the Cluster Randomized Trials***Description**

This function performs cluster randomized trials (CRT) analysis using model-robust standardization estimators to estimate the cluster-average and individual-average treatment effect. It handles different outcome mean models (GLM, LMM, GEE, GLMM) and supports both continuous, binary, and count outcomes with options for different correlation structures and scales (risk difference, risk ratio and odds ratio).

Usage

```
MRStdCRT_fit(
  formula,
  data,
  cluster,
  trt,
  trtprob = rep(0.5, nrow(data)),
  method,
  family = gaussian(link = "identity"),
  constr,
  scale,
  alpha = 0.05
)
```

Arguments

formula	A formula with a single untransformed response column and an intercept. Supported terms are untransformed covariate main effects and two-way interactions between treatment and one covariate; each interacting covariate must also appear as a main effect. The treatment main effect is added automatically. For example, $y \sim x + A:x$ and $y \sim A * x$ are supported when <code>trt = "A"</code> . Transformations, interactions between covariates, higher-order interactions, dot notation, no-intercept formulas, and term subtraction are not supported. Compute transformed covariates as separate data columns before specifying the formula.
data	A data frame where categorical variables should already be converted to dummy variables.
cluster	A string representing the column name of the cluster ID in the data frame. Cluster IDs must not be missing.
trt	A string representing the column name of the treatment assignment per cluster (0=control, 1=treatment). Values must be numeric or logical 0/1, without missing values, and constant within each cluster.
trtprob	A numeric vector of probabilities of assignment to treatment ($A = 1$), with one value per data row or per cluster. Values must be finite, strictly between 0 and

	1, and constant within each cluster. Row-level values must follow the order of data. Named cluster-level values are matched by cluster ID, and their names must uniquely and exactly match all cluster IDs; unnamed values follow the order of first appearance of clusters. If NULL, the common probability is estimated as the proportion of clusters assigned to treatment and re-estimated in each jackknife sample. The default is <code>rep(0.5, nrow(data))</code> .
method	A string specifying the outcome mean model. Possible values are: - 'GLM': generalized linear model on cluster-level means (binary/continuous outcome). - 'LMM': linear mixed model on individual-level observations (continuous outcome). - 'GEE': marginal models fitted by generalized estimating equations. - 'GLMM': generalized linear mixed model; Gaussian identity models are fitted by maximum likelihood as linear mixed models.
family	The link function for the outcome. Can be one of the following: - 'gaussian(link = "identity")': for continuous outcomes. Default is <code>gaussian("identity")</code> . - 'binomial(link = "logit")': for binary outcomes. - 'poisson(link = "log")': for count outcomes. - 'gaussian(link = "logit")': for binary outcomes with logit link to model the generalized linear model.
corstr	A string specifying the correlation structure for GEE models (e.g., "exchangeable", "independence").
scale	A string specifying the risk measure of interest. Can be 'RD' (risk difference), 'RR' (relative risk), or 'OR' (odds ratio).
alpha	A numeric value for the type-I error rate. Default is 0.05.

Value

A list with the following components: - 'estimate': A summary table of estimates. - 'm': Number of clusters. - 'N': Total number of observations per cluster. - 'family': The family used for the model. - 'model': The method used for the outcome mean model.

Examples

```
utils::data("ppact", package = "MRStdCRT")

fit <- MRStdCRT_fit(
  formula = PEGS ~ AGE + FEMALE + comorbid + Dep_OR_Anx + pain_count + PEGS_bl +
    BL_benzo_flag + BL_avg_daily + satisfied_primary + n,
  data     = ppact,
  cluster  = "CLUST",
  trt      = "INTERVENTION",
  trtprob  = NULL,
  method   = "GEE",
  corstr   = "independence",
  scale    = "RR"
)
summary(fit)
```

MRStdCRT_point

*Model-robust standardization in CRT Point Estimate***Description**

This function calculates a model-robust point estimate for a clustered randomized trial (CRT).

Usage

```
MRStdCRT_point(
  formula,
  data,
  cluster,
  trt,
  trtprob,
  family = gaussian(link = "identity"),
  constr,
  method = "GLM",
  scale
)
```

Arguments

formula	A formula with a single untransformed response column and an intercept. Supported terms are untransformed covariate main effects and two-way interactions between treatment and one covariate; each interacting covariate must also appear as a main effect. The treatment main effect is added automatically. For example, $y \sim x + A:x$ and $y \sim A * x$ are supported when <code>trt = "A"</code> . Transformations, interactions between covariates, higher-order interactions, dot notation, no-intercept formulas, and term subtraction are not supported. Compute transformed covariates as separate data columns before specifying the formula.
data	A data frame where categorical variables should already be converted to dummy variables.
cluster	A string representing the column name of the cluster ID in the data frame. Cluster IDs must not be missing.
trt	A string representing the column name of the treatment assignment per cluster (0=control, 1=treatment). Values must be numeric or logical 0/1, without missing values, and constant within each cluster.
trtprob	A numeric vector of probabilities of assignment to treatment ($A = 1$), with one value per individual in the same row order as data. Values must be finite, strictly between 0 and 1, and constant within each cluster.
family	The link function for the outcome. Can be one of the following: - 'gaussian(link = "identity")': for continuous outcomes. Default is <code>gaussian("identity")</code> - 'binomial(link = "logit")': for binary outcomes. - 'poisson(link = "log")': for count outcomes. - 'gaussian(link = "logit")': for binary outcomes with logit link to model the generalized linear model.

corstr	A string specifying the correlation structure for GEE models (e.g., "exchangeable", "independence").
method	A string specifying the outcome mean model. Possible values are: - 'GLM': Generalized linear model on cluster-level means (continuous/binary outcome). - 'LMM': linear mixed model on individual-level observations (continuous outcome). - 'GEE': marginal models fitted by generalized estimating equations. - 'GLMM': generalized linear mixed model.
scale	A string specifying the risk measure of interest. Can be 'RD' (risk difference), 'RR' (relative risk), or 'OR' (odds ratio).

Value

A list with the following components: - 'data1': A data frame containing all individual-level observations. - 'data_clus': A data frame containing all cluster-level summaries. - 'c(cate,iate,test_NICS)': A vector containing: (i) cate: point estimate for cluster-average treatment effect; (ii) iate: point estimate for individual-average treatment effect; (iii) test_NICS: value of test statistics for non-informative cluster sizes.

ppact

Example Dataset: PPACT

Description

The Pain Program of Active Coping and Training (PPACT) is a large-scale, mixed methods, cluster-randomized trial (CRT) to compare the effectiveness of an integrated, interdisciplinary program versus usual care in treating patients with chronic pain on long-term opioid treatment (CP-LOT). The primary outcome is the impact of pain (assessed using the PEGS)

Usage

ppact

Format

A data frame with primary outcome, cluster-level, individual level covariates:

SID Study ID

CLUST Cluster

INTERVENTION Study arm

AGE Patient age at randomization

FEMALE Participant gender

comorbid Diagnosis of 2 or more of the chronic medical conditions in 6 month prior to randomization

Dep_OR_Anx Anxiety and/or depression diagnosis in 6 months prior to randomization

pain_count Number of different pain types from which participants have diagnoses in 12 months prior to randomization

BL_benzo_flag Benzodiazepine dispensed in 6 months prior to randomization

BL_avg_daily Average morphine miligram equivalents dose per day in 6 month prior to randomization

PEGS_bl PEGS score at baseline

satisfied_primary Satisfaction with primary care services in prior 3 months

PEGS PEGS score

n cluster size

Source

ClinicalTrials.gov: NCT02113592, The manuscript of the study’s main outcomes is published in the Annals of Internal Medicine (<https://doi.org/10.7326/M21-1436>).

summary.MRS_obj	Summarize a MRS_obj Fit
-----------------	-------------------------

Description

Print a concise summary of a model-robust standardization CRT fit, including the c-ATE and i-ATE estimates with SEs and CIs.

Usage

```
## S3 method for class 'MRS_obj'
summary(object, ...)
```

Arguments

object An object of class MRS_obj, as returned by MRStdCRT_fit().

... Additional arguments (currently ignored).

Value

Invisibly returns the original MRS_obj object, after printing:

- Fitting method and family,
- Number of clusters and cluster sizes,
- A four-column table (Estimate, SE, CI, p-value) with rownames c-ATE and i-ATE. The confidence level is $100(1 - \alpha)\%$, using the alpha stored in the fit,
- The NICS test statistic and p-value.

Index

* **datasets**

data_sim_binary, [2](#)

data_sim_continuous, [3](#)

ppact, [7](#)

data_sim_binary, [2](#)

data_sim_continuous, [3](#)

MRStdCRT_fit, [4](#)

MRStdCRT_point, [6](#)

ppact, [7](#)

summary.MRS_obj, [8](#)