Package: MRStdCRT (via r-universe)

October 30, 2026

Type Package

Title Model-Robust Standardization in Cluster-Randomized Trials

Version 0.1.0

Author Jiaqi Tong, Changjun Li, Xi Fang, Chao Cheng, Bingkai Wang, Fan Li.

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

Description Implements model-robust standardization for cluster-randomized trials (CRTs). Provides functions that standardize user-specified regression models to estimate marginal treatment effects. The estimands include the cluster-average and individual-average treatment effects, with utilities for variance estimation and example simulation datasets.

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), stats

VignetteBuilder knitr

Suggests knitr, rmarkdown

RoxygenNote 7.3.2

NeedsCompilation no

Config/pak/sysreqs cmake make libicu-dev

Repository https://deckardt98.r-universe.dev

Date/Publication 2026-09-30 03:45:10 UTC

RemoteUrl https://github.com/deckardt98/mrstdcrt

RemoteRef HEAD

RemoteSha cb09b28d0093aaae55962a437fca84fc902e528d

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` *Example Dataset: Simulated CRT (binary outcome)*

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` *Example Dataset: Simulated CRT (continuous outcome)*

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"),
  corstr,
  scale,
  jack = 1,
  alpha = 0.05
)
```

Arguments

formula	A formula for the outcome mean model, including covariates.
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.
trt	A string representing the column name of the treatment assignment per cluster (0=control, 1=treatment).
trtprob	A vector of treatment probabilities per cluster (for each individual), conditional on covariates. Default is rep(0.5,nrow(data))
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.

family	The link function for the outcome. Can be one of the following: - ‘gaussian(link = "identity")‘: for continuous outcomes. Default is gaussian("identity"). - ‘binomial(link = "logit")‘: for binary outcomes. - ‘poisson(link = "log")‘: for count outcomes. - ‘gaussian(link = "logit")‘: for binary outcomes with logit link to model the genealized 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).
jack	A numeric value (1, 2, or 3) specifying the type of jackknife standard error estimate. Type 1 is the standard jackknife, and type 3 is recommended for small numbers of clusters. Default is 1.
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

```
## Not run:
ppact_prob <- ppact %>%
  group_by(CLUST) %>%
 mutate(first_trt = first(INTERVENTION)) %>%
 ungroup() %>%
 mutate(prob_A_1 = mean(first_trt == 1, na.rm = TRUE),  # Proportion trt=1
        prob_A_0 = mean(first_trt == 0, na.rm = TRUE)) %>%
 mutate(assigned_value = ifelse(INTERVENTION == 1, prob_A_1, prob_A_0))

prob <- ppact_prob$assigned_value

example <- 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 = prob,
 method = "GEE",
 corstr = "independence",
 scale = "RR"
)

## End(Not run)
```

MRStdCRT_point *Modet-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"),
  corstr,
  method = "GLM",
  scale
)
```

Arguments

formula	A formula for the outcome mean model, including covariates.
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.
trt	A string representing the column name of the treatment assignment per cluster.
trtprob	A vector of treatment probabilities per cluster (for each individual), conditional on covariates. Default is rep(0.5,nrow(data))
family	The link function for the outcome. Can be one of the following: - 'gaussian(link = "identity")': for continuous outcomes. Default is gaussian("identity") - 'binomial(link = "logit")': for binary outcomes. - 'poisson(link = "log")': for count outcomes. - 'gaussian(link = "logit")': for binary outcomes with logit link to model the genealized 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 (continous/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 contaning 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 three-column table (Estimate, SE, 95% CI) with rownames `c-ATE` and `i-ATE`,
- 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