Package: RCT2 (via r-universe)

Version 0.0.1

Date 2022-10-16

Title Designing and Analyzing Two-Stage Randomized Experiments

Maintainer Kosuke Imai <imai@harvard.edu>

Depends R (>= 3.1.0)

Description Provides various statistical methods for designing and
analyzing two-stage randomized controlled trials using the
methods developed by Imai, Jiang, and Malani (2021)
<doi:10.1080/01621459.2020.1775612> and (2022+)
<doi:10.48550/arXiv.2011.07677>. The package enables the
estimation of direct and spillover effects, conduct hypotheses
tests, and conduct sample size calculation for two-stage
randomized controlled trials.

License GPL (>= 2)

URL https://github.com/kosukeimai/RCT2

BugReports https://github.com/kosukeimai/RCT2/issues

RoxygenNote 7.2.1

Suggests testthat, knitr, rmarkdown

VignetteBuilder knitr

Imports sandwich, AER, stats, quadprog

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

RemoteUrl https://github.com/kosukeimai/rct2

RemoteRef HEAD

RemoteSha 6b5fcc7253bba0e242d175b37cf07b00ce373d7f

Contents

CADEparamreg	*Regression-based method for the ITT effects and the complier average direct effect/spillover effect*

Description

This function computes the point estimates and variance estimates of the direct effect and spillover effect for ITT and CADE/CASE

Usage

```
CADEparamreg(data, assign.prob, ci.level = 0.95)
```

Arguments

data	A data frame containing the relevant variables. The names for the variables should be: "Z" for the treatment assignment, "D" for the actual received treatment, "Y" for the outcome, "A" for the treatment assignment mechanism and "id" for the cluster ID. The variable for the cluster id should be a factor.
assign.prob	A double between 0 and 1 specifying the assignment probability to either assignment mechanism.
ci.level	A double between 0 and 1 specifying the confidence interval level to be output.

Details

For the details of the method implemented by this function, see the references.

Value

A list of class CADEparamreg which contains the following items:

ITT.DE	Estimate of direct effect under ITT regresion.
ITT.SE	Estimate of spillover effect under ITT regresion.
ITT.DE.CI	Confidence itnerval of direct effect under ITT regresion.
ITT.SE.CI	Confidence itnerval of spillover effect under ITT regresion.
IV.DE	Estimate of direct effect under IV regresion.
IV.SE	Estimate of spillover effect under IV regresion.

IV.DE.CI	Confidence interval of direct effect under IV regresion.
IV.SE.CI	Confidence interval of spillover effect under IV regresion.
IV.DE.CI	Confidence interval of direct effect under IV regresion.
ITT.tstat	t-stats from ITT regression.
IV.tstat	t-stats from IV regression.
ITT.pvals	p-values from ITT regression.
IV.pvals	p-values from IV regression.

data(india) india$id <- factor(india$id) CADEreg(india, ci.level = 0.90)

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Kosuke Imai, Zhichao Jiang and Anup Malani (2018). "Causal Inference with Interference and Noncompliance in the Two-Stage Randomized Experiments", *Technical Report*. Department of Politics, Princeton University.

CADErand	*Randomization-based method for the complier average direct effect and the complier average spillover effect*

Description

This function computes the point estimates and variance estimates of the complier average direct effect (CADE) and the complier average spillover effect (CASE). The estimators calculated using this function are either individual weighted or cluster-weighted. The point estimates and variances of ITT effects are also included.

Usage

```
CADErand(data, individual = 1, ci = 0.95)
```

Arguments

data	A data frame containing the relevant variables. The names for the variables should be: "Z" for the treatment assignment, "D" for the actual received treatment, "Y" for the outcome, "A" for the treatment assignment mechanism and "id" for the cluster ID. The variable for the cluster id should be a factor.
individual	A binary variable with TRUE for individual-weighted estimators and FALSE for cluster-weighted estimators.
ci	A numeric variable between 0 and 1 for the level of the confidence interval to be returned.

Details

For the details of the method implemented by this function, see the references.

Value

A list of class `CADErand` which contains the following items:

CADE	The point estimates of the CADE for each assignment mechanism.
CASE	The point estimate of CASE for each assignment mechanism.
var.CADE1	The variance estimate of CADE for each assignment mechanism.
var.CASE1	The variance estimate of CASE for each assignment mechanism.
DEY1	The point estimate of DEY for each assignment mechanism.
DED1	The point estimate of DED for each assignment mechanism.
var.DEY1	The variance estimate of DEY for each assignment mechanism.
var.DED1	The variance estimate of DED for each assignment mechanism.
SEY1	The point estimate of SEY for each pairwise groups of assignment mechanisms.
SED1	The point estimate of SED for each pairwise groups of assignment mechanisms.
var.SEY1	The variance estimate of SEY for each pairwise groups of assignment mechanisms.
var.SED1	The variance estimate of SED for each pairwise groups of assignment mechanisms.
lci.CADE	The left endpoint for the confidence intervals for the CADE from each assignment mechanism.
rci.CADE	The right endpoint for the confidence intervals for the CADE from each assignment mechanism.
lci.CASE	The left endpoint for the confidence intervals for the CASE from each assignment mechanism.
rci.CASE	The left endpoint for the confidence intervals for the CASE from each assignment mechanism.
lci.DEY	The left endpoint for the confidence intervals for the DEY from each assignment mechanism.
rci.DEY	The left endpoint for the confidence intervals for the DEY from each assignment mechanism.
lci.SEY	The left endpoint for the confidence intervals for the SEY from each pairwise groups of assignment mechanisms.
rci.SEY	The left endpoint for the confidence intervals for the SEY from each pairwise groups of assignment mechanism.
lci.DED	The left endpoint for the confidence intervals for the DED from each assignment mechanism.
rci.DED	The left endpoint for the confidence intervals for the DED from each assignment mechanism.
lci.SED	The left endpoint for the confidence intervals for the SED from each pairwise groups of assignment mechanism.
rci.SED	The left endpoint for the confidence intervals for the SED from each pairwise groups of assignment mechanism.

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Kosuke Imai, Zhichao Jiang and Anup Malani (2018). "Causal Inference with Interference and Noncompliance in the Two-Stage Randomized Experiments", *Technical Report*. Department of Politics, Princeton University.

Examples

```
data(india)
india$id <- factor(india$id)
CADErand(india, 0.95)
```

CADEreg | *Regression-based method for the complier average direct effect*

Description

This function computes the point estimates of the complier average direct effect (CADE) and four different variance estimates: the HC2 variance, the cluster-robust variance, the cluster-robust HC2 variance and the variance proposed in the reference. The estimators calculated using this function are cluster-weighted, i.e., the weights are equal for each cluster. To obtain the indivudal-weighted estimators, please multiply the recieved treatment and the outcome by n_jJ/N, where n_j is the number of individuals in cluster j, J is the number of clusters and N is the total number of individuals.

Usage

```
CADEreg(data, ci.level = 0.95)
```

Arguments

data | A data frame containing the relevant variables. The names for the variables should be: "Z" for the treatment assignment, "D" for the actual received treatment, "Y" for the outcome, "A" for the treatment assignment mechanism and "id" for the cluster ID. The variable for the cluster id should be a factor.

ci.level | A double between 0 and 1 specifying the confidence interval level to be output.

Details

For the details of the method implemented by this function, see the references.

Value

A list of class `CADEreg` which contains the following items:

`CADE1`	The point estimate of CADE(1).
`CADE0`	The point estimate of CADE(0).
`var1.clu`	The cluster-robust variance of CADE(1).
`var0.clu`	The cluster-robust variance of CADE(0).
`var1.clu.hc2`	The cluster-robust HC2 variance of CADE(1).
`var0.clu.hc2`	The cluster-robust HC2 variance of CADE(0).
`var1.hc2`	The HC2 variance of CADE(1).
`var0.hc2`	The HC2 variance of CADE(0).
`var1.ind`	The individual-robust variance of CADE(1).
`var0.ind`	The individual-robust variance of CADE(0).
`var1.reg`	The proposed variance of CADE(1).
`var0.reg`	The proposed variance of CADE(0).

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Kosuke Imai, Zhichao Jiang and Anup Malani (2018). "Causal Inference with Interference and Noncompliance in the Two-Stage Randomized Experiments", *Technical Report*. Department of Politics, Princeton University.

Examples

```
data(india)
india$id <- factor(india$id)
CADEreg(india, ci.level = 0.90)
```

CalAPO   *Point Estimation and Variance for the unit-level direct effect (ADE),*
                                 marginal direct effect (MDE), and unit level spillover effect (ASE)

Description

This function calculates the estimated average potential outcomes Y(z,a), point estimates for the ADE, MDE, and ASE, and conservative covariance matrix estimates.

Usage

```
CalAPO(data)
```

Arguments

data   A data frame containing the relevant variables. The names for the variables should be "Z" for the treatment assignment, "Y" for the treatment outcome, "A" for the treatment assignment mechanism, and "id" for the cluster ID. The variable for the cluster ID should be a factor.

Details

For the details of the method implemented by this function, see the references.

Value

A list of class `CalAPO` which contains the following items:

Y.hat	Estimate of the average potential outcomes.
ADE.est	Estimate of the unit level direct effect.
MDE.est	Estimate of the marginal direct effect.
ASE.est	Estimate of the unti level spillover effect.
cov.hat	Conservative covariance matrix for the estimated potential outcomes.
var.hat.ADE	Estimated variance of the ADE.
var.hat.MDE	Estimated variance of the MDE.
var.hat.ASE	Estimated variance of the ASE.

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Zhichao Jiang, Kosuke Imai (2020). "Statistical Inference and Power Analysis for Direct and Spillover Effects in Two-Stage Randomized Experiments", *Technical Report*.

Examples

```
data(jd)
data_LTFC <- data.frame(jd$assigned, jd$pct0, jd$cdd6m, jd$anonale)
colnames(data_LTFC) <- c("Z", "A", "Y", "id")
test <- CalAPO(data_LTFC)
print(CalAPO(data_LTFC))
```

---

calpara	*Sample size parameter calculations for detecting a specific alternative*

Description

This function calculates the parameters needed for the method to calculate sample size references.

Usage

```
calpara(data)
```

Arguments

data   A data frame containing the relevant variables. The names for the variables
              should be "Z" for the treatment assignment, "Y" for the treatment outcome,
              "A" for the treatment assignment mechanism, and "id" for the cluster ID. The
              variable for the cluster ID should be a factor.

Value

A list of class `calpara` which contains the following item:

sigmaw	The within-cluster variance of the potential outcomes, with the assumption that the all of the variances the same.
sigmab	The between-cluster variance of the potential outcomes, with the assumption that all of the variances are the same.
r	The intraclass correlation coefficient with respect to the potential outcomes.
sigma.tot	The total variance of the potential outcomes.
n.avg	The mean of the number of treated observations by cluster.

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Zhichao Jiang, Kosuke Imai (2020). "Statistical Inference and Power Analysis for Direct and Spillover Effects in Two-Stage Randomized Experiments", *Technical Report*.

Examples

```
data(jd)
data_LTFC <- data.frame(jd$assigned, jd$pct0, jd$cdd6m, jd$anonale)
colnames(data_LTFC) <- c("Z", "A", "Y", "id")
var.LTFC <- calpara(data_LTFC)
```

Calsamplesize	*Sample size calculations for detecting a specific alternative*

Description

This function calculates the sample size needed to detect a specific alternative hypothesis with a given power at a given significance level. For the details of the method implemented by this function, see the references.

Usage

```
Calsamplesize(data, mu, qa, alpha = 0.05, beta = 0.2)
```

Arguments

data	A data frame containing the relevant variables. The names for the variables should be "Z" for the treatment assignment, "Y" for the treatment outcome, "A" for the treatment assignment mechanism, and "id" for the cluster ID. The variable for the cluster ID should be a factor.
mu	The effect size (i.e. the largest direct effect across treatment assignment mechanisms).
qa	The proportions of different treatment assignment mechanisms.
alpha	The given significance level (default 0.05).
beta	The given power level (default 0.2).

Value

A list of class `sampleSRE` which contains the following item:

samplesize	A list of the calculated necessary nubmer of clusters for each assignment mechanism in order to detect a specific alternative with a given power at a given significance level.

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, `https://imai.fas.harvard.edu/`; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Zhichao Jiang, Kosuke Imai (2020). "Statistical Inference and Power Analysis for Direct and Spillover Effects in Two-Stage Randomized Experiments", *Technical Report*.

---

india	*Replication Data for: Causal Inference with Interference and Non-compliance in Two-Stage Randomized Experiments.*

Description

Replication Data for: Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments.

Usage

```
data(india)
```

Format

A data frame with columns:

id The id for the village.

DistrictId The id for the district.

Z The treatment status for the individual.

A The treatment assignment mechanism.

D Whether or not the individual enrolled.

Y The hospital expenditure.

X Enumeration of the patients.

Source

[doi:10.7910/DVN/N7D9LS](doi:10.7910/DVN/N7D9LS)

| jd | *Replication Data for: Statistical Inference and Power Analysis for Direct and Spillover Effects in Two-Stage Randomized Experiments* |

Description

Replication Data for: Statistical Inference and Power Analysis for Direct and Spillover Effects in Two-Stage Randomized Experiments

Usage

```
data(jd)
```

Format

A data frame with columns:

anonale The local employment agency.

tempsc_av Categorical variable for full-time work at time of assignment (1: 1-4 months, 2: 4-8 months, 3: 8-12 months, 4: 12+ months)

assigned An indicator variable for whether or not the individual is assigned to treatment.

pct0 The share of the local population treated (as a decimal).

cdi An indicator variable for whether the individual works on a permanent contract 8 months after assignment.

cdd6m An indicator variable for whether the individual works in CDD (LTFC-time contract) for more than 6 months, 8 months after the assignment.

emploidur An indicator variable for whether the individual works on a permanent or LTFC-term contract for more than 6 months, 8 months after the assignment.

tempsc An indicator variable for whether the individual works full time, 8 months after the assignment.

salaire The individual's salary in Euros.

| print.regression | *Print Method for the RCT2 Package* |

Description

This function prints a nicely formatted summary of the three functions in the RCT2 package.

Usage

```
## S3 method for class 'regression'
print(x, ...)
```

Arguments

x   A list object generated by running one of the analyses on a data set.

...   ignored

Details

For the details of the method implemented by this function, see the references.

Value

NULL

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Kosuke Imai, Zhichao Jiang and Anup Malani (2018). "Causal Inference with Interference and Noncompliance in the Two-Stage Randomized Experiments", *Technical Report*. Department of Politics, Princeton University.

Test2SRE   *Hypothesis testing for three null hypotheses*

Description

This function tests the null hypotheses of no direct effect, no marginal direct effect, and no spillover effect.

Usage

```
Test2SRE(data, effect = "DE", alpha = 0.05)
```

Arguments

data   A data frame containing the relevant variables. The names for the variables should be "Z" for the treatment assignment, "Y" for the treatment outcome, "A" for the treatment assignment mechanism, and "id" for the cluster ID. The variable for the cluster ID should be a factor.

effect   Specify which null hypothesis to be tested. "DE" for direct effect, "ME" for marginal effect, and "SE" for spillover effect.

alpha   The level of significance at which the test is to be run (default is 0.05).

Details

For the details of the method implemented by this function, see the references.

Value

A list of class `Test2SRE` which contains the following item:

rej   Rejection region for test conducted.

Author(s)

Kosuke Imai, Department of Statistics, Harvard University <imai@harvard.edu>, https://imai.fas.harvard.edu/; Zhichao Jiang, School of Public Health and Health Sciences, University of Massachusetts Amherst <zhichaojiang@umass.edu>; Karissa Huang, Department of Statistics, Harvard College <krhuang@college.harvard.edu>

References

Zhichao Jiang, Kosuke Imai (2020). "Statistical Inference and Power Analysis for Direct and Spillover Effects in Two-Stage Randomized Experiments", *Technical Report*.

Examples

```
data(jd)
data_LTFC <- data.frame(jd$assigned, jd$pct0, jd$cdd6m, jd$anonale)
colnames(data_LTFC) <- c("Z", "A", "Y", "id")
Test2SRE(data_LTFC, effect="MDE", alpha=0.05)
```

Index