

Package ‘sreg’

August 25, 2025

Type Package

Title Stratified Randomized Experiments

Version 2.0.2

Description Estimate average treatment effects (ATEs) in stratified randomized experiments. `sreg` supports a wide range of stratification designs, including matched pairs, n-tuple designs, and larger strata with many units — possibly of unequal size across strata. 'sreg' is designed to accommodate scenarios with multiple treatments and cluster-level treatment assignments, and accommodates optimal linear covariate adjustment based on baseline observable characteristics. 'sreg' computes estimators and standard errors based on Bugni, Canay, Shaikh (2018) <doi:10.1080/01621459.2017.1375934>; Bugni, Canay, Shaikh, Tabord-Meehan (2024+) <doi:10.48550/arXiv.2204.08356>; Jiang, Linton, Tang, Zhang (2023+) <doi:10.48550/arXiv.2201.13004>; Bai, Jiang, Romano, Shaikh, and Zhang (2024) <doi:10.1016/j.jeconom.2024.105740>; Bai (2022) <doi:10.1257/aer.20201856>; Bai, Romano, and Shaikh (2022) <doi:10.1080/01621459.2021.1883437>; Liu (2024+) <doi:10.48550/arXiv.2301.09016>; and Cybaum (2024) <doi:10.3982/QE2475>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Imports dplyr, tidyr, purrr, extraDistr, rlang, cli, ggplot2, viridis

Suggests haven, knitr, rmarkdown, testthat

Depends R (>= 2.10)

Config/testthat/edition 3

VignetteBuilder knitr

URL <https://github.com/jutrifonov/sreg>

BugReports <https://github.com/jutrifonov/sreg/issues>

NeedsCompilation no

Author Juri Trifonov [aut, cre, cph],
Yuehao Bai [aut],
Azeem Shaikh [aut],
Max Tabord-Meehan [aut]

Maintainer Juri Trifonov <jutrifonov@u.northwestern.edu>
Repository CRAN
Date/Publication 2025-08-25 13:50:02 UTC

Contents

AEJapp	2
plot.sreg	3
print.sreg	4
sreg	5
sreg.rgen	8
Index	10

AEJapp	<i>Replication data for: Iron Deficiency and Schooling Attainment in Peru (Chong et al, 2016)</i>
--------	---

Description

The data is taken from Chong et al. (2016), who study the effect of iron deficiency anemia (i.e., anemia caused by a lack of iron) on school-age children’s educational attainment and cognitive ability in Peru.

Usage

```
data("AEJapp")
```

Format

A data frame with 215 observations on the 62 variables.

Source

Chong, A., Cohen, I., Field, E., Nakasone, E., and Torero, M. (2016). Replication data for: Iron Deficiency and Schooling Attainment in Peru. Nashville, TN: American Economic Association [publisher], 2016. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-12. [doi:10.3886/E113624V1](https://doi.org/10.3886/E113624V1).

References

Chong, A., Cohen, I., Field, E., Nakasone, E., and Torero, M. (2016). Iron Deficiency and Schooling Attainment in Peru. *American Economic Journal: Applied Economics*, 8(4), 222–255. [doi:10.1257/app.20140494](https://doi.org/10.1257/app.20140494).

Examples

```
data(AEJapp)
```

plot.sreg

*Plot Method for 'sreg' Objects***Description**

Visualize estimated ATEs and confidence intervals for objects of class sreg.

Usage

```
## S3 method for class 'sreg'
plot(
  x,
  treatment_labels = NULL,
  title = "Estimated ATEs with Confidence Intervals",
  bar_fill = NULL,
  point_shape = 23,
  point_size = 3,
  point_fill = "white",
  point_stroke = 1.2,
  point_color = "black",
  label_color = "black",
  label_size = 4,
  bg_color = NULL,
  grid = TRUE,
  zero_line = TRUE,
  y_axis_title = NULL,
  x_axis_title = NULL,
  ...
)
```

Arguments

x	An object of class sreg.
treatment_labels	Optional vector of treatment labels to display on the y-axis. If NULL, default labels like "Treatment 1", "Treatment 2", etc., are used.
title	Optional plot title. Defaults to "Estimated ATEs with Confidence Intervals".
bar_fill	Optional fill color(s) for the confidence interval bars. Can be NULL (default viridis scale), a single color, or a vector of two colors for a gradient.
point_shape	Optional shape of the point used to mark the estimated ATE. Default is 23 (a diamond).
point_size	Optional size of the point marking the ATE.
point_fill	Optional fill color of the ATE point shape.
point_stroke	Optional stroke (border) thickness of the ATE point shape.
point_color	Optional outline color of the ATE point.

label_color	Optional color of the text label displaying the estimate and standard error.
label_size	Optional size of the text label displaying the estimate and standard error.
bg_color	Optional background color of the plot panel. If NULL, the default theme background is used.
grid	Optional logical flag. If TRUE (default), grid lines are shown; if FALSE, they are removed.
zero_line	Optional logical flag. If TRUE (default), a vertical dashed line at 0 is added for reference.
y_axis_title	Optional title of the y-axis. If NULL, no y-axis label is added.
x_axis_title	Optional title of the x-axis. If NULL, no x-axis label is added.
...	Additional arguments passed to other methods.

Value

Invisibly returns the ggplot object. Called for its side effects (i.e., generating a plot).

print.sreg	<i>Print sreg Objects</i>
------------	---------------------------

Description

Print the summary table of estimation results for sreg objects.

Usage

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

Arguments

x	An object of class sreg.
...	Additional arguments passed to other methods.

Value

No return value, called for side effects.

Examples

```
data <- sreg.rgen(n = 200, tau.vec = c(0.1), n.strata = 4, cluster = TRUE)
Y <- data$Y
S <- data$S
D <- data$D
X <- data.frame("x_1" = data$x_1, "x_2" = data$x_2)
result <- sreg(Y, S, D, G.id = NULL, Ng = NULL, X)
print(result)
```

sreg	<i>Estimate Average Treatment Effects (ATEs) and Corresponding Standard Errors</i>
------	--

Description

Estimate the ATE(s) and the corresponding standard error(s) for a (collection of) treatment(s) relative to a control.

Usage

```
sreg(
  Y,
  S = NULL,
  D,
  G.id = NULL,
  Ng = NULL,
  X = NULL,
  HC1 = TRUE,
  small.strata = FALSE
)
```

Arguments

Y	a numeric $n \times 1$ vector/matrix/data.frame/tibble of the observed outcomes
S	a numeric $n \times 1$ vector/matrix/data.frame/tibble of strata indicators indexed by $\{1, 2, 3, \dots\}$; if NULL then the estimation is performed assuming no stratification
D	a numeric $n \times 1$ vector/matrix/data.frame/tibble of treatments indexed by $\{0, 1, 2, \dots\}$, where $D = 0$ denotes the control
G.id	a numeric $n \times 1$ vector/matrix/data.frame/tibble of cluster indicators; if NULL then estimation is performed assuming treatment is assigned at the individual level
Ng	a numeric $n \times 1$ vector/matrix/data.frame/tibble of cluster sizes; if NULL then Ng is assumed to be equal to the number of available observations in every cluster
X	a matrix/data.frame/tibble with columns representing the covariate values for every observation; if NULL then the estimator without linear adjustments is applied. (Note: sreg cannot use individual-level covariates for covariate adjustment in cluster-randomized experiments. Any individual-level covariates will be aggregated to their cluster-level averages)
HC1	a TRUE/FALSE logical argument indicating whether the small sample correction should be applied to the variance estimator
small.strata	a TRUE/FALSE logical argument indicating whether the estimators for small strata (i.e., strata with few units, such as matched pairs or n-tuples) should be used.

Value

An object of class `sreg` that is a list containing the following elements:

- `tau.hat`: a $1 \times |\mathcal{A}|$ vector of ATE estimates, where $|\mathcal{A}|$ represents the number of treatments
- `se.rob`: a $1 \times |\mathcal{A}|$ vector of standard errors estimates, where $|\mathcal{A}|$ represents the number of treatments
- `t.stat`: a $1 \times |\mathcal{A}|$ vector of t -statistics, where $|\mathcal{A}|$ represents the number of treatments
- `p.value`: a $1 \times |\mathcal{A}|$ vector of corresponding p -values, where $|\mathcal{A}|$ represents the number of treatments
- `CI.left`: a $1 \times |\mathcal{A}|$ vector of the left bounds of the 95% as. confidence interval
- `CI.right`: a $1 \times |\mathcal{A}|$ vector of the right bounds of the 95% as. confidence interval
- `data`: an original data of the form `data.frame(Y, S, D, G.id, Ng, X)`
- `lin.adj`: a `data.frame` representing the covariates that were used in implementing linear adjustments
- `small.strata`: a TRUE/FALSE logical argument indicating whether the estimators for small strata (e.g., matched pairs or n-tuples) were used
- `HC1`: a TRUE/FALSE logical argument indicating whether the small sample correction (HC1) was applied to the variance estimator

Author(s)**Authors:**

Juri Trifonov <jutrifonov@u.northwestern.edu>

Yuehao Bai <yuehao.bai@usc.edu>

Azeem Shaikh <amshaikh@uchicago.edu>

Max Tabord-Meehan <m.tabordmeehan@utoronto.ca>

Maintainer:

Juri Trifonov <jutrifonov@u.northwestern.edu>

References

- Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018). Inference Under Covariate-Adaptive Randomization. *Journal of the American Statistical Association*, 113(524), 1784–1796, doi:10.1080/01621459.2017.1375934.
- Bugni, F., Canay, I., Shaikh, A., and Tabord-Meehan, M. (2024+). Inference for Cluster Randomized Experiments with Non-ignorable Cluster Sizes. *Forthcoming in the Journal of Political Economy: Microeconomics*, doi:10.48550/arXiv.2204.08356.
- Jiang, L., Linton, O. B., Tang, H., and Zhang, Y. (2023+). Improving Estimation Efficiency via Regression-Adjustment in Covariate-Adaptive Randomizations with Imperfect Compliance. *Forthcoming in Review of Economics and Statistics*, doi:10.48550/arXiv.2204.08356.
- Bai, Y., Jiang, L., Romano, J. P., Shaikh, A. M., and Zhang, Y. (2024). Covariate adjustment in experiments with matched pairs. *Journal of Econometrics*, 241(1), doi:10.1016/j.jeconom.2024.105740.

- Bai, Y. (2022). Optimality of Matched-Pair Designs in Randomized Controlled Trials. *American Economic Review*, 112(12), doi:10.1257/aer.20201856.
- Bai, Y., Romano, J. P., and Shaikh, A. M. (2022). Inference in Experiments With Matched Pairs. *Journal of the American Statistical Association*, 117(540), doi:10.1080/01621459.2021.1883437.
- Liu, J. (2024). Inference for Two-stage Experiments under Covariate-Adaptive Randomization. doi:10.48550/arXiv.2301.09016.
- Cytrynbaum, M. (2024). Covariate Adjustment in Stratified Experiments. *Quantitative Economics*, 15(4), 971–998, doi:10.3982/QE2475.

Examples

```
library("sreg")
library("dplyr")
library("haven")
### Example 1. Simulated Data.
data <- sreg.rgen(n = 1000, tau.vec = c(0), n.strata = 4, cluster = FALSE)
Y <- data$Y
S <- data$S
D <- data$D
X <- data.frame("x_1" = data$x_1, "x_2" = data$x_2)
result <- sreg(Y, S, D, G.id = NULL, Ng = NULL, X)
print(result)
### Example 2. Empirical Data.
?AEJapp
data("AEJapp")
data <- AEJapp
head(data)
Y <- data$gradesq34
D <- data$treatment
S <- data$class_level
data.clean <- data.frame(Y, D, S)
data.clean <- data.clean %>%
  mutate(D = ifelse(D == 3, 0, D))
Y <- data.clean$Y
D <- data.clean$D
S <- data.clean$S
table(D = data.clean$D, S = data.clean$S)
result <- sreg(Y, S, D)
print(result)
pills <- data$pills_taken
age <- data$age_months
data.clean <- data.frame(Y, D, S, pills, age)
data.clean <- data.clean %>%
  mutate(D = ifelse(D == 3, 0, D))
Y <- data.clean$Y
D <- data.clean$D
S <- data.clean$S
X <- data.frame("pills" = data.clean$pills, "age" = data.clean$age)
result <- sreg(Y, S, D, G.id = NULL, X = X)
print(result)
### Example 3. Matched Pairs (small strata).
```

```

data <- sreg.rgen(
  n = 1000, tau.vec = c(1.2), cluster = FALSE,
  small.strata = TRUE, k = 2, treat.sizes = c(1, 1)
)
Y <- data$Y
S <- data$S
D <- data$D
X <- data.frame("x_1" = data$x_1, "x_2" = data$x_2)
result <- sreg(Y = Y, S = S, D = D, X = X, small.strata = TRUE)
print(result)

```

sreg.rgen

Generate a Pseudo-Random Sample under the Stratified Block Randomization Design

Description

The function generates the observed outcomes, treatment assignments, strata indicators, cluster indicators, cluster sizes, and covariates for estimating the treatment effect within the context of a stratified block randomization design under the covariate-adaptive randomization (CAR).

Usage

```

sreg.rgen(
  n,
  Nmax = 50,
  n.strata = 10,
  tau.vec = c(0),
  gamma.vec = c(0.4, 0.2, 1),
  cluster = TRUE,
  is.cov = TRUE,
  small.strata = FALSE,
  k = 3,
  treat.sizes = c(1, 1, 1)
)

```

Arguments

n	a total number of observations in a sample
Nmax	a maximum size of generated clusters (maximum number of observations in a cluster)
n.strata	an integer specifying the number of strata
tau.vec	a numeric $1 \times \mathcal{A} $ vector of treatment effects, where $ \mathcal{A} $ represents the number of treatments
gamma.vec	a numeric 1×3 vector of parameters corresponding to covariates
cluster	a TRUE/FALSE argument indicating whether the dgp should use a cluster-level treatment assignment or individual-level

<code>is.cov</code>	a TRUE/FALSE argument indicating whether the dgp should include covariates or not
<code>small.strata</code>	a TRUE/FALSE argument indicating whether the data-generating process should use a small-strata design (e.g., matched pairs, n-tuples)
<code>k</code>	an integer specifying the number of units per stratum when <code>small.strata = TRUE</code>
<code>treat.sizes</code>	a numeric $1 \times (\mathcal{A} + 1)$ vector specifying the number of units assigned to each treatment within a stratum; the first element corresponds to control units ($D = 0$), the second to the first treatment ($D = 1$), and so on

Value

An object that is a 'data.frame' with n observations containing the generated values of the following variables:

- `Y`: a numeric $n \times 1$ vector of observed outcomes
- `S`: a numeric $n \times 1$ vector of strata indicators
- `D`: a numeric $n \times 1$ vector of treatments indexed by $\{0, 1, 2, \dots\}$, where $D = 0$ denotes the control
- `G.id`: a numeric $n \times 1$ vector of cluster indicators
- `X`: a data.frame with columns representing the covariate values for every observation

Examples

```
data <- sreg.rgen(n = 1000, tau.vec = c(0), n.strata = 4, cluster = TRUE)
```

Index

* **datasets**

AEJapp, [2](#)

AEJapp, [2](#)

plot.sreg, [3](#)

print.sreg, [4](#)

sreg, [5](#)

sreg.rgen, [8](#)