Package 'ciccr'

October 29, 2020

Type Package

Title Causal Inference in Case-Control Studies

Version 0.2.0

Description Estimation and inference methods for causal relative and attributable risk in case-control and case-population studies.
Semiparametrically efficient estimation of the aggregated (log) odds ratio and causal inference procedures for relative and attributable risk.
For more details, see the paper by Jun and Lee (2020), ``Causal Inference in Case-Control Studies,'' <arXiv:2004.08318 [econ.EM]>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Imports stats

Suggests knitr, rmarkdown, testthat, MASS

VignetteBuilder knitr

Depends R (>= 2.10)

URL https://github.com/sokbae/ciccr/

BugReports https://github.com/sokbae/ciccr/issues

NeedsCompilation no

Author Sung Jae Jun [aut],
Sokbae Lee [aut, cre]

Maintainer Sokbae Lee <sl3841@columbia.edu>

Repository CRAN

Date/Publication 2020-10-29 10:00:02 UTC

R topics documented:

ACS_CC 2
ACS_CP 3
avg_AR_logit 3
avg_RR_logit 5
ciccr 6
cicc_AR 6
cicc_plot 8
cicc_RR 9
trim_pr 10

Index **11**

ACS_CC *ACS_CC*

Description

A case-control sample extracted from American Community Survey (ACS) 2018, restricted to white males residing in California with at least a bachelor's degree. The original ACS dataset is not from case-control sampling, but this case-control sample is obtained by the following procedure. The case sample is composed of 921 individuals whose income is top-coded. The control sample of equal size is randomly drawn without replacement from the pool of individuals whose income is not top-coded. Age is restricted to be between 25 and 70.

Usage

```
ACS_CC
```

Format

A data frame with 1842 rows and 4 variables:

age age, in years

ind industry code, in four digits

baplus 1 if BA or higher; 0 otherwise

topincome 1 if income is top-coded; 0 otherwise

Source

https://usa.ipums.org/usa/

ACS_CP *ACS_CP*

Description

A case-population sample extracted from American Community Survey (ACS) 2018, restricted to white males residing in California with at least a bachelor's degree. The original ACS dataset is not from case-population sampling, but this case-population sample is obtained by the following procedure. The case sample is composed of 921 individuals whose income is top-coded. The control sample of equal size is randomly drawn with replacement from all observations and its top-coded status is coded missing. Age is restricted to be between 25 and 70.

Usage

```
ACS_CP
```

Format

A data frame with 1842 rows and 4 variables:

age age, in years

ind industry code, in four digits

baplus 1 if BA or higher; 0 otherwise

topincome 1 if an observation belongs to the case sample; NA otherwise

Source

https://usa.ipums.org/usa/

avg_AR_logit *An Average of the Upper Bound on Causal Attributable Risk*

Description

Averages the upper bound on causal attributable risk using prospective and retrospective logistic regression models under the monotone treatment response (MTR) and monotone treatment selection (MTS) assumptions.

Usage

```
avg_AR_logit(
  y,
  t,
  x,
  sampling = "cc",
  p_upper = 1L,
  length = 21L,
  interaction = TRUE,
  eps = 1e-08
)
```

Arguments

y	n-dimensional vector of binary outcomes
t	n-dimensional vector of binary treatments
x	n by d matrix of covariates
sampling	'cc' for case-control sampling; 'cp' for case-population sampling (default = 'cc')
p_upper	specified upper bound for the unknown true case probability (default = 1)
length	specified length of a sequence from 0 to p_upper (default = 21)
interaction	TRUE if there are interaction terms in the retrospective logistic model; FALSE if not (default = TRUE)
eps	a small constant that determines the trimming of the estimated probabilities. Specifically, the estimate probability is trimmed to be between eps and 1-eps (default = 1e-8).

Value

An S3 object of type "ciccr". The object has the following elements.

est	(length)-dimensional vector of the average of the upper bound of causal attributable risk
pseq	(length)-dimensional vector of a grid from 0 to p_upper

References

Jun, S.J. and Lee, S. (2020). Causal Inference in Case-Control Studies. https://arxiv.org/abs/2004.08318.

Manski, C.F. (1997). Monotone Treatment Response. Econometrica, 65(6), 1311-1334.

Manski, C.F. and Pepper, J.V. (2000). Monotone Instrumental Variables: With an Application to the Returns to Schooling. Econometrica, 68(4), 997-1010.

Examples

```
# use the ACS_CC dataset included in the package.
  y = ciccr::ACS_CC$topincome
  t = ciccr::ACS_CC$baplus
  x = ciccr::ACS_CC$age
  results = avg_AR_logit(y, t, x, sampling = 'cc')
```

---

avg_RR_logit *An Average of the Log Odds Ratio*

Description

Averages the log odds ratio using retrospective logistic regression.

Usage

```
avg_RR_logit(y, t, x, w = "control")
```

Arguments

y	n-dimensional vector of binary outcomes
t	n-dimensional vector of binary treatments
x	n by d matrix of covariates
w	'case' if the average is conditional on the case sample; 'control' if it is conditional on the control sample default w = 'control'

Value

An S3 object of type "ciccr". The object has the following elements.

est	a scalar estimate of the weighted average of the log odds ratio using retrospective logistic regression
se	standard error

References

Jun, S.J. and Lee, S. (2020). Causal Inference in Case-Control Studies. https://arxiv.org/abs/2004.08318.

Examples

```
# use the ACS_CC dataset included in the package
  y = ciccr::ACS_CC$topincome
  t = ciccr::ACS_CC$baplus
  x = ciccr::ACS_CC$age
# use 'case' to condition on the distribution of covariates given y = 1
  results = avg_RR_logit(y, t, x, 'case')
```

ciccr *ciccr: a package for causal inference in case-control and case-population studies*

Description

The ciccr package provides methods for causal inference in case-control and case-population studies under the monotone treatment response (MTR) and monotone treatment selection (MTS) assumptions.

Functions

The package includes the following:

- cicc_plot: plots upper bounds on relative and attributable risk.
- cicc_RR: carries out causal inference on relative risk.
- avg_RR_logit: averages the log odds ratio using retrospective logistic regression.
- cicc_AR: carries out causal inference on attributable risk.
- avg_AR_logit: averages the upper bound on causal attributable risk using prospective and retrospective logistic regression models.
- ACS_CC: provides an illustrative case-control sample.
- ACS_CP: provides an illustrative case-population sample.

References

Jun, S.J. and Lee, S. (2020). Causal Inference in Case-Control Studies. https://arxiv.org/abs/2004.08318.

Manski, C.F. (1997). Monotone Treatment Response. Econometrica, 65(6), 1311-1334.

Manski, C.F. and Pepper, J.V. (2000). Monotone Instrumental Variables: With an Application to the Returns to Schooling. Econometrica, 68(4), 997-1010.

cicc_AR *Causal Inference on Attributable Risk*

Description

Provides an upper bound on the average of attributable risk under the monotone treatment response (MTR) and monotone treatment selection (MTS) assumptions.

Usage

```
cicc_AR(
  y,
  t,
  x,
  sampling = "cc",
  p_upper = 1L,
  cov_prob = 0.95,
  length = 21L,
  interaction = TRUE,
  no_boot = 0L,
  eps = 1e-08
)
```

Arguments

y	n-dimensional vector of binary outcomes
t	n-dimensional vector of binary treatments
x	n by d matrix of covariates
sampling	'cc' for case-control sampling; 'cp' for case-population sampling (default = 'cc')
p_upper	a specified upper bound for the unknown true case probability (default = 1)
cov_prob	coverage probability of a confidence interval (default = 0.95)
length	specified length of a sequence from 0 to p_upper (default = 21)
interaction	TRUE if there are interaction terms in the retrospective logistic model; FALSE if not (default = TRUE)
no_boot	number of bootstrap repetitions to compute the confidence intervals (default = 0)
eps	a small constant that determines the trimming of the estimated probabilities. Specifically, the estimate probability is trimmed to be between eps and 1-eps (default = 1e-8).

Value

An S3 object of type "ciccr". The object has the following elements:

est	(length)-dimensional vector of the upper bounds on the average of attributable risk
ci	(length)-dimensional vector of the upper ends of pointwise one-sided confidence intervals
pseq	(length)-dimensional vector of a grid from 0 to p_upper
cov_prob	the nominal coverage probability
return_code	status of existence of missing values in bootstrap replications

References

Jun, S.J. and Lee, S. (2020). Causal Inference in Case-Control Studies. https://arxiv.org/abs/2004.08318.

Manski, C.F. (1997). Monotone Treatment Response. Econometrica, 65(6), 1311-1334.

Manski, C.F. and Pepper, J.V. (2000). Monotone Instrumental Variables: With an Application to the Returns to Schooling. Econometrica, 68(4), 997-1010.

Examples

```
# use the ACS_CC dataset included in the package.
  y = ciccr::ACS_CC$topincome
  t = ciccr::ACS_CC$baplus
  x = ciccr::ACS_CC$age
  results_AR = cicc_AR(y, t, x, sampling = 'cc', no_boot = 100)
```

---

cicc_plot *Plotting Upper Bounds on Relative and Attributable Risk*

Description

Plots upper bounds on relative and attributable risk

Usage

```
cicc_plot(
  results,
  parameter = "RR",
  sampling = "cc",
  save_plots = FALSE,
  file_name = Sys.Date(),
  plots_ctl = 0.3,
  plots_dir = FALSE
)
```

Arguments

results	estimation results from either cicc_RR or cicc_AR
parameter	'RR' for relative risk; 'AR' for attributable risk (default = 'RR')
sampling	'cc' for case-control sampling; 'cp' for case-population sampling (default = 'cc')
save_plots	TRUE if the plots are saved as pdf files; FALSE if not (default = FALSE)
file_name	the pdf file name to save the plots (default = Sys.Date())
plots_ctl	value to determine the topleft position of the legend in the figure a large value makes the legend far away from the confidence intervals (default = 0.3)
plots_dir	a directory where the plots are saved (default = FALSE); plots will be saved under "(current working directory)/figures" by default.

Value

A X-Y plot where the X axis shows the range of p from 0 to p_upper and the Y axis depicts both point estimates and the upper end point of the one-sided confidence intervals.

References

Jun, S.J. and Lee, S. (2020). Causal Inference in Case-Control Studies. https://arxiv.org/abs/2004.08318.

Examples

```
# use the ACS_CC dataset included in the package.
  y = ciccr::ACS_CC$topincome
  t = ciccr::ACS_CC$baplus
  x = ciccr::ACS_CC$age
  results = cicc_RR(y, t, x)
  cicc_plot(results)
```

---

cicc_RR *Causal Inference on Relative Risk*

Description

Provides upper bounds on the average of log relative risk under the monotone treatment response (MTR) and monotone treatment selection (MTS) assumptions.

Usage

```
cicc_RR(y, t, x, sampling = "cc", cov_prob = 0.95)
```

Arguments

y	n-dimensional vector of binary outcomes
t	n-dimensional vector of binary treatments
x	n by d matrix of covariates
sampling	'cc' for case-control sampling; 'cp' for case-population sampling (default = 'cc')
cov_prob	coverage probability of a uniform confidence band (default = 0.95)

Value

An S3 object of type "ciccr". The object has the following elements:

est	estimates of the upper bounds on the average of log relative risk at p=0 and p=1
se	pointwise standard errors at p=0 and p=1
ci	the upper end points of the uniform confidence band at p=0 and p=1
pseq	two end points: p=0 and p=1

References

Jun, S.J. and Lee, S. (2020). Causal Inference in Case-Control Studies. https://arxiv.org/abs/2004.08318.

Manski, C.F. (1997). Monotone Treatment Response. Econometrica, 65(6), 1311-1334.

Manski, C.F. and Pepper, J.V. (2000). Monotone Instrumental Variables: With an Application to the Returns to Schooling. Econometrica, 68(4), 997-1010.

Examples

```
# use the ACS_CC dataset included in the package.
  y = ciccr::ACS_CC$topincome
  t = ciccr::ACS_CC$baplus
  x = ciccr::ACS_CC$age
  results_RR = cicc_RR(y, t, x, sampling = 'cc', cov_prob = 0.95)
```

---

trim_pr *Trimming the estimates to be strictly between 0 and 1*

Description

Trimming the estimates to be strictly between 0 and 1

Usage

```
trim_pr(ps, eps = 1e-08)
```

Arguments

ps	n-dimensional vector of estimated probabilities
eps	a small constant that determines the trimming of the estimated probabilities. Specifically, the estimate probability is trimmed to be between eps and 1-eps (default = 1e-8).

Value

ps_tr	n-dimensional trimmed estimates

Index

∗ **datasets**
ACS_CC, 2
ACS_CP, 3

ACS_CC, 2, *6*
ACS_CP, 3, *6*
avg_AR_logit, 3, *6*
avg_RR_logit, 5, *6*

cicc_AR, *6*, 6
cicc_plot, *6*, 8
cicc_RR, *6*, 9
ciccr, 6

trim_pr, 10