Package 'aihuman'

May 7, 2026

Type Package

Title Experimental Evaluation of Algorithm-Assisted Human Decision-Making

Version 1.0.1

Date 2025-5-2

Description Provides statistical methods for analyzing experimental evaluation of the causal impacts of algorithmic recommendations on human decisions developed by Imai, Jiang, Greiner, Halen, and Shin (2023) <doi:10.1093/jrsssa/qnad010> and Ben-Michael, Greiner, Huang, Imai, Jiang, and Shin (2024) <doi:10.48550/arXiv.2403.12108>. The data used for this paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

License GPL (>= 2)

URL https://github.com/sooahnshin/aihuman

BugReports https://github.com/sooahnshin/aihuman/issues

Encoding UTF-8

RoxygenNote 7.3.2

Imports Rcpp, coda, stats, magrittr, purrr, abind, foreach, parallel, doParallel, ggplot2, dplyr, tidyr, metR, MASS, GLMMadaptive, gbm, tidyselect, stringr, forcats

LinkingTo Rcpp, RcppArmadillo, RcppEigen

Depends R (>= 4.1.0)

Suggests knitr, rmarkdown

VignetteBuilder knitr

LazyData true

NeedsCompilation yes

Author Sooahn Shin [aut, cre] (ORCID: <https://orcid.org/0000-0001-6213-2197>), Zhichao Jiang [aut], Kosuke Imai [aut]

Maintainer Sooahn Shin <sooahnshin@g.harvard.edu>

Repository CRAN

Date/Publication 2025-05-07 15:20:02 UTC

Contents

`aihuman-package` *Experimental Evaluation of Algorithm-Assisted Human Decision-Making*

Description

Provides statistical methods for analyzing experimental evaluation of the causal impacts of algorithmic recommendations on human decisions developed by Imai, Jiang, Greiner, Halen, and Shin (2023) <doi:10.1093/jrsssa/qnad010> and Ben-Michael, Greiner, Huang, Imai, Jiang, and Shin (2024) <doi:10.48550/arXiv.2403.12108>. The data used for this paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

Package Content

Index of help topics:

```
APCEsummary             Summary of APCE
APCEsummaryipw          Summary of APCE for frequentist analysis
AiEvalmcmc              Gibbs sampler for the main analysis
BootstrapAPCEipw        Bootstrap for estimating variance of APCE
BootstrapAPCEipwRE      Bootstrap for estimating variance of APCE with
                        random effects
BootstrapAPCEipwREparallel
                        Bootstrap for estimating variance of APCE with
                        random effects
CalAPCE                 Calculate APCE
CalAPCEipw              Compute APCE using frequentist analysis
```

CalAPCEipwRE	Compute APCE using frequentist analysis with random effects
CalAPCEparallel	Calculate APCE using parallel computing
CalDIM	Calculate diff-in-means estimates
CalDIMsubgroup	Calculate diff-in-means estimates
CalDelta	Calculate the delta given the principal stratum
CalFairness	Calculate the principal fairness
CalOptimalDecision	Calculate optimal decision & utility
CalPS	Calculate the proportion of principal strata (R)
FTAdata	Interim Dane data with failure to appear (FTA) as an outcome
HearingDate	Interim court event hearing date
NCAdata	Interim Dane data with new criminal activity (NCA) as an outcome
NVCAdata	Interim Dane data with new violent criminal activity (NVCA) as an outcome
PSAdata	Interim Dane PSA data
PlotAPCE	Plot APCE
PlotDIMdecisions	Plot diff-in-means estimates
PlotDIMoutcomes	Plot diff-in-means estimates
PlotFairness	Plot the principal fairness
PlotOptimalDecision	Plot optimal decision
PlotPS	Plot the proportion of principal strata (R)
PlotSpilloverCRT	Plot conditional randomization test
PlotSpilloverCRTpower	Plot power analysis of conditional randomization test
PlotStackedBar	Stacked barplot for the distribution of the decision given psa
PlotStackedBarDMF	Stacked barplot for the distribution of the decision given DMF recommendation
PlotUtilityDiff	Plot utility difference
PlotUtilityDiffCI	Plot utility difference with 95 interval
SpilloverCRT	Conduct conditional randomization test
SpilloverCRTpower	Conduct power analysis of conditional randomization test
TestMonotonicity	Test monotonicity
TestMonotonicityRE	Test monotonicity with random effects
aihuman-package	Experimental Evaluation of Algorithm-Assisted Human Decision-Making
compute_bounds_aipw	Compute Risk (AI v. Human)
compute_nuisance_functions	Fit outcome/decision and propensity score models
compute_nuisance_functions_ai	Fit outcome/decision and propensity score models conditioning on the AI recommendation

compute_stats	Compute Risk (Human+AI v. Human)
compute_stats_agreement	Agreement of Human and AI Decision Makers
compute_stats_aipw	Compute Risk (Human+AI v. Human)
compute_stats_subgroup	Compute Risk (Human+AI v. Human) for a Subgroup Defined by AI Recommendation
crossfit	Crossfitting for nuisance functions
g_legend	Pulling ggplot legend
hearingdate_synth	Synthetic court event hearing date
plot_agreement	Visualize Agreement
plot_diff_ai_aipw	Visualize Difference in Risk (AI v. Human)
plot_diff_human	Visualize Difference in Risk (Human+AI v. Human)
plot_diff_human_aipw	Visualize Difference in Risk (Human+AI v. Human)
plot_diff_subgroup	Visualize Difference in Risk (Human+AI v. Human) for a Subgroup Defined by AI Recommendation
plot_preference	Visualize Preference
psa_synth	Synthetic PSA data
synth	Synthetic data
table_agreement	Table of Agreement

Maintainer

Sooahn Shin <sooahnshin@g.harvard.edu>

Author(s)

Sooahn Shin [aut, cre] (<https://orcid.org/0000-0001-6213-2197>), Zhichao Jiang [aut], Kosuke Imai [aut]

`AiEvalmcmc` *Gibbs sampler for the main analysis*

Description

See Appendix S5 for more details.

Usage

```
AiEvalmcmc(
  data,
  rho = 0,
  Sigma0.beta.inv = NULL,
  Sigma0.alpha.inv = NULL,
```

```
  sigma0 = NULL,
  beta = NULL,
  alpha = NULL,
  theta = NULL,
  delta = NULL,
  n.mcmc = 5 * 10,
  verbose = FALSE,
  out.length = 10,
  beta.zx.off = FALSE,
  theta.z.off = FALSE
)
```

Arguments

data	A data.frame or matrix of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
rho	A sensitivity parameter. The default is 0 which implies the unconfoundedness assumption (Assumption 4).
Sigma0.beta.inv	Inverse of the prior covariance matrix of beta. The default is a diagonal matrix with 0.01 diagonal entries.
Sigma0.alpha.inv	Inverse of the prior covariance matrix of alpha. The default is a diagonal matrix with 0.01 diagonal entries.
sigma0	Prior variance of the cutoff points (theta and delta)
beta	Initial value for beta.
alpha	Initial value for alpha.
theta	Initial value for theta.
delta	Initial value for delta.
n.mcmc	The total number of MCMC iterations. The default is 50.
verbose	A logical argument specified to print the progress on the screen. The default is FALSE.
out.length	An integer to specify the progress on the screen. If verbose = TRUE, every out.length-th iteration is printed on the screen. The default is 10.
beta.zx.off	A logical argument specified to exclude the interaction terms (Z by X) from the model. The default is FALSE.
theta.z.off	A logical argument specified to set same cutoffs theta for treatment and control group. The default is FALSE.

Value

An object of class mcmc containing the posterior samples.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 2)
```

`APCEsummary` *Summary of APCE*

Description

Summary of average principal causal effects (APCE) with ordinal decision.

Usage

```
APCEsummary(apce.mcmc)
```

Arguments

`apce.mcmc`	APCE.mcmc array generated from `CalAPCE` or `CalAPCEparallel`.

Value

A `data.frame` that consists of mean and quantiles (2.5

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_apce <- CalAPCE(data = synth, mcmc.re = sample_mcmc, subgroup = subgroup_synth)
sample_apce_summary <- APCEsummary(sample_apce[["APCE.mcmc"]])
```

`APCEsummaryipw` *Summary of APCE for frequentist analysis*

Description

Summary of average principal causal effects (APCE) with ordinal decision with frequentist results.

Usage

```
APCEsummaryipw(
  G1_est,
  G2_est,
  G3_est,
  G4_est,
  G5_est,
  G1_boot,
  G2_boot,
  G3_boot,
  G4_boot,
  G5_boot,
  name.group = c("Overall", "Female", "Male", "Non-white\nMale", "White\nMale")
)
```

Arguments

`G1_est`	List generated from `CalAPCEipw` for the first subgroup.
`G2_est`	List generated from `CalAPCEipw` for the second subgroup.
`G3_est`	List generated from `CalAPCEipw` for the third subgroup.
`G4_est`	List generated from `CalAPCEipw` for the fourth subgroup.
`G5_est`	List generated from `CalAPCEipw` for the fifth subgroup.
`G1_boot`	List generated from `BootstrapAPCEipw` for the first subgroup.
`G2_boot`	List generated from `BootstrapAPCEipw` for the second subgroup.
`G3_boot`	List generated from `BootstrapAPCEipw` for the third subgroup.
`G4_boot`	List generated from `BootstrapAPCEipw` for the fourth subgroup.
`G5_boot`	List generated from `BootstrapAPCEipw` for the fifth subgroup.
`name.group`	A list of character vectors for the label of five subgroups.

Value

A `data.frame` that consists of mean and quantiles (2.5

Examples

```
data(synth)
synth$SexWhite <- synth$Sex * synth$White
freq_apce <- CalAPCEipw(synth)
boot_apce <- BootstrapAPCEipw(synth, rep = 10)
# subgroup analysis
data_s0 <- subset(synth, synth$Sex == 0, select = -c(Sex, SexWhite))
freq_s0 <- CalAPCEipw(data_s0)
boot_s0 <- BootstrapAPCEipw(data_s0, rep = 10)
data_s1 <- subset(synth, synth$Sex == 1, select = -c(Sex, SexWhite))
freq_s1 <- CalAPCEipw(data_s1)
boot_s1 <- BootstrapAPCEipw(data_s1, rep = 10)
data_s1w0 <- subset(synth, synth$Sex == 1 & synth$White == 0, select = -c(Sex, White, SexWhite))
freq_s1w0 <- CalAPCEipw(data_s1w0)
boot_s1w0 <- BootstrapAPCEipw(data_s1w0, rep = 10)
data_s1w1 <- subset(synth, synth$Sex == 1 & synth$White == 1, select = -c(Sex, White, SexWhite))
freq_s1w1 <- CalAPCEipw(data_s1w1)
boot_s1w1 <- BootstrapAPCEipw(data_s1w1, rep = 10)

freq_apce_summary <- APCEsummaryipw(
  freq_apce, freq_s0, freq_s1, freq_s1w0, freq_s1w1,
  boot_apce, boot_s0, boot_s1, boot_s1w0, boot_s1w0
)
PlotAPCE(freq_apce_summary,
  y.max = 0.25, decision.labels = c(
    "signature", "small cash",
    "middle cash", "large cash"
  ), shape.values = c(16, 17, 15, 18),
  col.values = c("blue", "black", "red", "brown", "purple"), label = FALSE
)
```

`BootstrapAPCEipw` *Bootstrap for estimating variance of APCE*

Description

Estimate variance of APCE for frequentist analysis using bootstrap. See S7 for more details.

Usage

```
BootstrapAPCEipw(data, rep = 1000)
```

Arguments

`data`	A `data.frame` or `matrix` of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
`rep`	Size of bootstrap

Value

An object of class `list` with the following elements:

`P.D1.boot`	An array with dimension rep by (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is rep (size of bootstrap), dimension 2 is (k+1) values of D from 0 to k, dimension 3 is (k+2) values of R from 0 to k+1.
`P.D0.boot`	An array with dimension rep by (k+1) by (k+2) for quantity P(D(0)=d| R=r).
`APCE.boot`	An array with dimension rep by (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).
`P.R.boot`	An array with dimension rep by (k+2) for quantity P(R=r) for r from 0 to (k+1).
`alpha.boot`	An array with estimated alpha for each bootstrap.
`delta.boot`	An array with estimated delta for each bootstrap.

Examples

```
data(synth)
set.seed(123)
boot_apce <- BootstrapAPCEipw(synth, rep = 100)
```

`BootstrapAPCEipwRE` *Bootstrap for estimating variance of APCE with random effects*

Description

Estimate variance of APCE for frequentist analysis with random effects using bootstrap. See S7 for more details.

Usage

```
BootstrapAPCEipwRE(
  data,
  rep = 1000,
  fixed,
  random,
  CourtEvent_HearingDate,
  nAGQ = 1
)
```

Arguments

`data`	A `data.frame` or `matrix` of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
`rep`	Size of bootstrap

`fixed` A formula for the fixed-effects part of the model to fit.

`random` A formula for the random-effects part of the model to fit.

`CourtEvent_HearingDate`
The court event hearing date.

`nAGQ` Integer scalar - the number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood. Defaults to 1, corresponding to the Laplace approximation.

Value

An object of class `list` with the following elements:

`P.D1.boot` An array with dimension rep by (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is rep (size of bootstrap), dimension 2 is (k+1) values of D from 0 to k, dimension 3 is (k+2) values of R from 0 to k+1.

`P.D0.boot` An array with dimension rep by (k+1) by (k+2) for quantity P(D(0)=d| R=r).

`APCE.boot` An array with dimension rep by (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).

`P.R.boot` An array with dimension rep by (k+2) for quantity P(R=r) for r from 0 to (k+1).

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
data(hearingdate_synth)
synth$CourtEvent_HearingDate <- hearingdate_synth
set.seed(123)
boot_apce_re <- BootstrapAPCEipwRE(synth,
  fixed = "Y ~ Sex + White + Age +
                  CurrentViolentOffense + PendingChargeAtTimeOfOffense +
                  PriorMisdemeanorConviction + PriorFelonyConviction +
                  PriorViolentConviction + D",
  random = "~ 1|CourtEvent_HearingDate"
)
```

`BootstrapAPCEipwREparallel` *Bootstrap for estimating variance of APCE with random effects*

Description

Estimate variance of APCE for frequentist analysis with random effects using bootstrap. See S7 for more details.

Usage

```
BootstrapAPCEipwREparallel(data, rep = 1000, fixed, random, nAGQ = 1, size = 5)
```

Arguments

`data`	A `data.frame` or `matrix` of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
`rep`	Size of bootstrap
`fixed`	A formula for the fixed-effects part of the model to fit.
`random`	A formula for the random-effects part of the model to fit.
`nAGQ`	Integer scalar - the number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood. Defaults to 1, corresponding to the Laplace approximation.
`size`	The number of parallel computing. The default is `5`.

Value

An object of class `list` with the following elements:

`P.D1.boot`	An array with dimension rep by (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is rep (size of bootstrap), dimension 2 is (k+1) values of D from 0 to k, dimension 3 is (k+2) values of R from 0 to k+1.
`P.D0.boot`	An array with dimension rep by (k+1) by (k+2) for quantity P(D(0)=d| R=r).
`APCE.boot`	An array with dimension rep by (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).
`P.R.boot`	An array with dimension rep by (k+2) for quantity P(R=r) for r from 0 to (k+1).

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
data(hearingdate_synth)
synth$CourtEvent_HearingDate <- hearingdate_synth
set.seed(123)
boot_apce_re <- BootstrapAPCEipwREparallel(synth,
  fixed = "Y ~ Sex + White + Age +
                  CurrentViolentOffense + PendingChargeAtTimeOfOffense +
                  PriorMisdemeanorConviction + PriorFelonyConviction +
                  PriorViolentConviction + D",
  random = "~ 1|CourtEvent_HearingDate",
  size = 1
) # adjust the size
```

`CalAPCE` *Calculate APCE*

Description

Calculate average principal causal effects (APCE) with ordinal decision. See Section 3.4 for more details.

Usage

```
CalAPCE(
  data,
  mcmc.re,
  subgroup,
  name.group = c("overall", "Sex0", "Sex1", "Sex1 White0", "Sex1 White1"),
  rho = 0,
  burnin = 0,
  out.length = 500,
  c0 = 0,
  c1 = 0,
  ZX = NULL,
  save.individual.optimal.decision = FALSE,
  parallel = FALSE,
  optimal.decision.only = FALSE,
  dmf = NULL,
  fair.dmf.only = FALSE
)
```

Arguments

`data`	A `data.frame` or `matrix` of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y).

The column names of the latter three should be specified as "Z", "D", and "Y" respectively.

`mcmc.re` A `mcmc` object generated by `AiEvalmcmc()` function.

`subgroup` A list of numeric vectors for the index of each of the five subgroups.

`name.group` A list of character vectors for the label of five subgroups.

`rho` A sensitivity parameter. The default is `0` which implies the unconfoundedness assumption (Assumption 4).

`burnin` A proportion of burnin for the Markov chain. The default is `0`.

`out.length` An integer to specify the progress on the screen. Every `out.length`-th iteration is printed on the screen. The default is `500`.

`c0` The cost of an outcome. See Section 3.7 for more details. The default is `0`.

`c1` The cost of an unnecessarily harsh decision. See Section 3.7 for more details. The default is `0`.

`ZX` The data matrix for interaction terms. The default is the interaction between Z and all of the pre-treatment covariates (X).

`save.individual.optimal.decision`
A logical argument specified to save individual optimal decision rules. The default is `FALSE`.

`parallel` A logical argument specifying whether parallel computing is conducted. Do not change this argument manually.

`optimal.decision.only`
A logical argument specified to compute only the optimal decision rule. The default is `FALSE`.

`dmf` A numeric vector of binary DMF recommendations. If `null`, use judge's decisions (0 if the decision is 0 and 1 o.w; e.g., signature or cash bond).

`fair.dmf.only` A logical argument specified to compute only the fairness of given DMF recommendations. The default is `FALSE`. Not used in the analysis for the JRSSA paper.

Value

An object of class `list` with the following elements:

`P.D1.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is each posterior sample; dimension 2 is subgroup, dimension 3 is (k+1) values of D from 0 to k, dimension 4 is (k+2) values of R from 0 to k+1.

`P.D0.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for quantity P(D(0)=d| R=r).

`APCE.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).

`P.R.mcmc` An array with dimension n.mcmc by 5 by (k+2) for quantity P(R=r) for r from 0 to (k+1).

`Optimal.Z.mcmc` An array with dimension n.mcmc by 5 for the proportion of the cases where treatment (PSA provided) is optimal.

`Optimal.D.mcmc` An array with dimension n.mcmc by 5 by (k+1) for the proportion of optimal decision rule (average over observations). If `save.individual.optimal.decision = TRUE`, the dimension would be n by (k+1) (average over mcmc samples).

`P.DMF.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for the proportion of binary DMF recommendations. Not used in the analysis for the JRSSA paper.

`Utility.g_d.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n for the individual utility of judge's decisions.

`Utility.g_dmf.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n for the individual utility of DMF recommendation.

`Utility.diff.control.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n.mcmc for estimated difference in utility between judge's decisions and DMF recommendation among control group.

`Utility.diff.treated.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n.mcmc for estimated difference in utility between judge's decisions and DMF recommendation among treated group.

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 2)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_apce <- CalAPCE(data = synth, mcmc.re = sample_mcmc, subgroup = subgroup_synth)
```

CalAPCEipw *Compute APCE using frequentist analysis*

Description

Estimate propensity score and use Hajek estimator to compute APCE. See S7 for more details.

Usage

```
CalAPCEipw(data)
```

Arguments

data	A data.frame or matrix of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.

Value

An object of class list with the following elements:

P.D1	An array with dimension (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is (k+1) values of D from 0 to k, dimension 2 is (k+2) values of R from 0 to k+1.
P.D0	An array with dimension (k+1) by (k+2) for quantity P(D(0)=d| R=r).
APCE	An array with dimension (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).
P.R	An array with dimension (k+2) for quantity P(R=r) for r from 0 to (k+1).
alpha	An array with estimated alpha.
delta	An array with estimated delta.

Examples

```
data(synth)
freq_apce <- CalAPCEipw(synth)
```

CalAPCEipwRE *Compute APCE using frequentist analysis with random effects*

Description

Estimate propensity score and use Hajek estimator to compute APCE. See S7 for more details.

Usage

```
CalAPCEipwRE(data, fixed, random, nAGQ = 1)
```

Arguments

data	A data.frame or matrix of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
fixed	A formula for the fixed-effects part of the model to fit.
random	A formula for the random-effects part of the model to fit.
nAGQ	Integer scalar - the number of points per axis for evaluating the adaptive Gauss-Hermite approximation to the log-likelihood. Defaults to 1, corresponding to the Laplace approximation.

Value

An object of class list with the following elements:

P.D1	An array with dimension (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is (k+1) values of D from 0 to k, dimension 2 is (k+2) values of R from 0 to k+1.
P.D0	An array with dimension (k+1) by (k+2) for quantity P(D(0)=d| R=r).
APCE	An array with dimension (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).
P.R	An array with dimension (k+2) for quantity P(R=r) for r from 0 to (k+1).
alpha	An array with estimated alpha.
delta	An array with estimated delta.

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
data(hearingdate_synth)
synth$CourtEvent_HearingDate <- hearingdate_synth
freq_apce_re <- CalAPCEipwRE(synth,
  fixed = "Y ~ Sex + White + Age +
                  CurrentViolentOffense + PendingChargeAtTimeOfOffense +
                  PriorMisdemeanorConviction + PriorFelonyConviction +
                  PriorViolentConviction + D",
  random = "~ 1|CourtEvent_HearingDate"
)
```

`CalAPCEparallel` *Calculate APCE using parallel computing*

Description

Calculate average principal causal effects (APCE) with ordinal decision using parallel computing. See Section 3.4 for more details.

Usage

```
CalAPCEparallel(
  data,
  mcmc.re,
  subgroup,
  name.group = c("overall", "Sex0", "Sex1", "Sex1 White0", "Sex1 White1"),
  rho = 0,
  burnin = 0,
  out.length = 500,
  c0 = 0,
  c1 = 0,
  ZX = NULL,
  save.individual.optimal.decision = FALSE,
  optimal.decision.only = FALSE,
  dmf = NULL,
  fair.dmf.only = FALSE,
  size = 5
)
```

Arguments

`data`	A `data.frame` or `matrix` of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
`mcmc.re`	A `mcmc` object generated by `AiEvalmcmc()` function.
`subgroup`	A list of numeric vectors for the index of each of the five subgroups.
`name.group`	A list of character vectors for the label of five subgroups.
`rho`	A sensitivity parameter. The default is `0` which implies the unconfoundedness assumption (Assumption 4).
`burnin`	A proportion of burnin for the Markov chain. The default is `0`.
`out.length`	An integer to specify the progress on the screen. Every `out.length`-th iteration is printed on the screen. The default is `500`.
`c0`	The cost of an outcome. See Section 3.7 for more details. The default is `0`.
`c1`	The cost of an unnecessarily harsh decision. See Section 3.7 for more details. The default is `0`.

`ZX` The data matrix for interaction terms. The default is the interaction between Z and all of the pre-treatment covariates (X).

`save.individual.optimal.decision`
A logical argument specified to save individual optimal decision rules. The default is `FALSE`.

`optimal.decision.only`
A logical argument specified to compute only the optimal decision rule. The default is `FALSE`.

`dmf` A numeric vector of binary DMF recommendations. If `null`, use judge's decisions (0 if the decision is 0 and 1 o.w; e.g., signature or cash bond).

`fair.dmf.only` A logical argument specified to compute only the fairness of given DMF recommendations. The default is `FALSE`. Not used in the analysis for the JRSSA paper.

`size` The number of parallel computing. The default is `5`.

Value

An object of class `list` with the following elements:

`P.D1.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for quantity P(D(1)=d| R=r), dimension 1 is each posterior sample; dimension 2 is subgroup, dimension 3 is (k+1) values of D from 0 to k, dimension 4 is (k+2) values of R from 0 to k+1.

`P.D0.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for quantity P(D(0)=d| R=r).

`APCE.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for quantity P(D(1)=d| R=r)-P(D(0)=d| R=r).

`P.R.mcmc` An array with dimension n.mcmc by 5 by (k+2) for quantity P(R=r) for r from 0 to (k+1).

`Optimal.Z.mcmc` An array with dimension n.mcmc by 5 for the proportion of the cases where treatment (PSA provided) is optimal.

`Optimal.D.mcmc` An array with dimension n.mcmc by 5 by (k+1) for the proportion of optimal decision rule.

`P.DMF.mcmc` An array with dimension n.mcmc by 5 by (k+1) by (k+2) for the proportion of binary DMF recommendations. Not used in the analysis for the JRSSA paper.

`Utility.g_d.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n for the individual utility of judge's decisions.

`Utility.g_dmf.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n for the individual utility of DMF recommendation.

`Utility.diff.control.mcmc`
Included if `save.individual.optimal.decision = TRUE`. An array with dimension n.mcmc for estimated difference in utility between judge's decisions and DMF recommendation among control group.

`Utility.diff.treated.mcmc`
: Included if `save.individual.optimal.decision = TRUE`. An array with dimension n.mcmc for estimated difference in utility between judge's decisions and DMF recommendation among treated group.

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_apce <- CalAPCEparallel(
  data = synth, mcmc.re = sample_mcmc,
  subgroup = subgroup_synth,
  size = 1
) # adjust the size
```

CalDelta *Calculate the delta given the principal stratum*

Description

Calculate the maximal deviation of decisions probability among the distributions for different groups (delta) given the principal stratum (R).

Usage

```
CalDelta(r, k, pd0, pd1, attr)
```

Arguments

`r`	The given principal stratum.
`k`	The maximum decision (e.g., largest bail amount).
`pd0`	P.D0.mcmc array generated from `CalAPCE` or `CalAPCEparallel`.
`pd1`	P.D1.mcmc array generated from `CalAPCE` or `CalAPCEparallel`.
`attr`	The index of subgroups (within the output of CalAPCE/CalAPCEparallel) that corresponds to the protected attributes.

Value

A data.frame of the delta.

Examples

```
data(synth)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
sample_apce <- CalAPCE(
  data = synth, mcmc.re = sample_mcmc, subgroup = subgroup_synth,
  burnin = 0
)
CalDelta(0, 3, sample_apce[["P.D0.mcmc"]], sample_apce[["P.D1.mcmc"]], c(2, 3))
```

CalDIM *Calculate diff-in-means estimates*

Description

Calculate average causal effect based on diff-in-means estimator.

Usage

```
CalDIM(data)
```

Arguments

data	A data.frame of which columns includes a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y).

Value

A data.frame of diff-in-means estimates effect for each value of D and Y.

Examples

```
data(synth)
CalDIM(synth)
```

CalDIMsubgroup *Calculate diff-in-means estimates*

Description

Calculate average causal effect based on diff-in-means estimator.

Usage

```
CalDIMsubgroup(
  data,
  subgroup,
  name.group = c("Overall", "Female", "Male", "Non-white\nMale", "White\nMale")
)
```

Arguments

data	A data.frame of which columns includes a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y).
subgroup	A list of numeric vectors for the index of each of the five subgroups.
name.group	A character vector including the labels of five subgroups.

Value

A data.frame of diff-in-means estimates for each value of D and Y for each subgroup.

Examples

```
data(synth)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
CalDIMsubgroup(synth, subgroup = subgroup_synth)
```

CalFairness *Calculate the principal fairness*

Description

See Section 3.6 for more details.

Usage

```
CalFairness(apce, attr = c(2, 3))
```

Arguments

apce	The list generated from CalAPCE or CalAPCEparallel.
attr	The index of subgroups (within the output of CalAPCE/CalAPCEparallel) that corresponds to the protected attributes.

Value

A data.frame of the delta.

Examples

```
data(synth)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
sample_apce <- CalAPCE(
  data = synth, mcmc.re = sample_mcmc, subgroup = subgroup_synth,
  burnin = 0
)
CalFairness(sample_apce)
```

CalOptimalDecision *Calculate optimal decision & utility*

Description

(1) Calculate optimal decision for each observation given each of c0 (cost of an outcome) and c1 (cost of an unnecessarily harsh decision) from the lists. (2) Calculate difference in the expected utility between binary version of judge's decisions and DMF recommendations given each of c0 (cost of an outcome) and c1 (cost of an unnecessarily harsh decision) from the lists.

Usage

```
CalOptimalDecision(
  data,
  mcmc.re,
  c0.ls,
  c1.ls,
  dmf = NULL,
  rho = 0,
  burnin = 0,
  out.length = 500,
  ZX = NULL,
  size = 5,
```

```
  include.utility.diff.mcmc = FALSE
)
```

Arguments

data	A data.frame or matrix of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
mcmc.re	A mcmc object generated by AiEvalmcmc() function.
c0.ls	The list of cost of an outcome. See Section 3.7 for more details.
c1.ls	The list of cost of an unnecessarily harsh decision. See Section 3.7 for more details.
dmf	A numeric vector of binary DMF recommendations. If null, use judge's decisions (0 if the decision is 0 and 1 o.w; e.g., signature or cash bond).
rho	A sensitivity parameter. The default is 0 which implies the unconfoundedness assumption (Assumption 4).
burnin	A proportion of burnin for the Markov chain. The default is 0.
out.length	An integer to specify the progress on the screen. Every out.length-th iteration is printed on the screen. The default is 500.
ZX	The data matrix for interaction terms. The default is the interaction between Z and all of the pre-treatment covariates (X).
size	The number of parallel computing. The default is 5.
include.utility.diff.mcmc	A logical argument specifying whether to save Utility.diff.control.mcmc and Utility.diff.treated.mcmc for Figure S17. The default is FALSE.

Value

A data.frame of (1) the probability that the optimal decision for each observation being d in (0,1,...,k), (2) expected utility of binary version of judge's decision (g_d), (3) expected utility of binary DMF recommendation, and (4) the difference between (2) and (3). If include.utility.diff.mcmc = TRUE, returns a list of such data.frame and a data.frame that includes the result for mean and quantile of Utility.diff.control.mcmc and Utility.diff.treated.mcmc across mcmc samples.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
sample_optd <- CalOptimalDecision(
  data = synth, mcmc.re = sample_mcmc,
  c0.ls = seq(0, 5, 1), c1.ls = seq(0, 5, 1),
  size = 1
) # adjust the size
```

CalPS *Calculate the proportion of principal strata (R)*

Description

Calculate the proportion of each principal stratum (R).

Usage

```
CalPS(
  p.r.mcmc,
  name.group = c("Overall", "Female", "Male", "Non-white\nMale", "White\nMale")
)
```

Arguments

p.r.mcmc	P.R.mcmc array generated from CalAPCE or CalAPCEparallel.
name.group	A character vector including the labels of five subgroups.

Value

A data.frame of the proportion of each principal stratum.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_apce <- CalAPCE(
  data = synth, mcmc.re = sample_mcmc,
  subgroup = subgroup_synth
)
CalPS(sample_apce[["P.R.mcmc"]])
```

`compute_bounds_aipw` *Compute Risk (AI v. Human)*

Description

Compute the difference in risk between AI and human decision makers using AIPW estimators.

Usage

```
compute_bounds_aipw(
  Y,
  A,
  D,
  Z,
  X = NULL,
  nuis_funcs,
  nuis_funcs_ai,
  true.pscore = NULL,
  l01 = 1
)
```

Arguments

`Y`	An observed outcome (binary: numeric vector of 0 or 1).
`A`	An observed AI recommendation (binary: numeric vector of 0 or 1).
`D`	An observed decision (binary: numeric vector of 0 or 1).
`Z`	A treatment indicator (binary: numeric vector of 0 or 1).
`X`	Pretreatment covariate used for subgroup analysis (vector). Must be the same length as Y, D, Z, and A if provided. Default is NULL.
`nuis_funcs`	output from `compute_nuisance_functions`
`nuis_funcs_ai`	output from `compute_nuisance_functions_ai`
`true.pscore`	A vector of true propensity scores (numeric), if available. Optional.
`l01`	Ratio of the loss between false positives and false negatives

Value

A tibble the following columns:

- `Z_focal`: The focal treatment indicator. ‘1‘ indicates the treatment group.
- `Z_compare`: The comparison treatment indicator. ‘0‘ indicates the control group.
- `X`: Pretreatment covariate (if provided).
- `fn_diff_lb`: The lower bound of difference in false negatives
- `fn_diff_ub`: The upper bound of difference in false negatives

- `fp_diff_lb`: The lower bound of difference in false positives
- `fp_diff_ub`: The upper bound of difference in false positives
- `loss_diff_lb`: The lower bound of difference in loss
- `loss_diff_ub`: The upper bound of difference in loss
- `fn_diff_lb_se`: The standard error of the difference in false negatives
- `fn_diff_ub_se`: The standard error of the difference in false negatives
- `fp_diff_lb_se`: The standard error of the difference in false positives
- `fp_diff_ub_se`: The standard error of the difference in false positives
- `loss_diff_lb_se`: The standard error of the difference in loss
- `loss_diff_ub_se`: The standard error of the difference in loss

Examples

```
compute_bounds_aipw(
  Y = NCAdata$Y,
  A = PSAdata$DMF,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  nuis_funcs = nuis_func,
  nuis_funcs_ai = nuis_func_ai,
  true.pscore = rep(0.5, nrow(NCAdata)),
  X = NULL,
  l01 = 1
)
```

---

compute_nuisance_functions

Fit outcome/decision and propensity score models

Description

Fit (1) the decision model $m^D(z, X_i) := \Pr(D = 1 \mid Z = z, X = X_i)$ and (2) the outcome model $m^Y(z, X_i) := \Pr(Y = 1 \mid D = 0, Z = z, X = X_i)$ for each treatment group $z \in \{0, 1\}$ and (3) the propensity score model $e(1, X_i) := \Pr(Z = 1 \mid X = X_i)$.

Usage

```
compute_nuisance_functions(
  Y,
  D,
  Z,
  V,
  d_form = D ~ .,
  y_form = Y ~ .,
  ps_form = Z ~ .,
```

```
  distribution = "bernoulli",
  n.trees = 1000,
  shrinkage = 0.01,
  interaction.depth = 1,
  ...
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
V	Pretreatment covariates for nuisance functions. A vector, a matrix, or a data frame.
d_form	A formula for decision model where the dependent variable is D.
y_form	A formula for outcome model where the dependent variable is Y.
ps_form	A formula for propensity score model.
distribution	A distribution argument used in gbm function. Default is "bernoulli".
n.trees	Integer specifying the total number of trees to fit used in gbm function.
shrinkage	A shrinkage parameter used in gbm function.
interaction.depth	Integer specifying the maximum depth of each tree used in gbm function.
...	Additional arguments to be passed to gbm function called in crossfit

Value

A list with the following components:

z_models A data.frame with the following columns:

- idx Index of observation.
- d_pred Predicted probability of decision.
- y_pred Predicted probability of outcome.
- Z Treatment group.

pscore A vector of predicted propensity scores.

Examples

```
compute_nuisance_functions(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  V = NCAdata[, c("Sex", "White", "Age")],
  shrinkage = 0.01,
  n.trees = 1000
)
```

`compute_nuisance_functions_ai` *Fit outcome/decision and propensity score models conditioning on the AI recommendation*

Description

Fit (1) the decision model $m^D(z, a, X_i) := \Pr(D = 1 \mid Z = z, A = a, X = X_i)$ and (2) the outcome model $m^Y(z, a, X_i) := \Pr(Y = 1 \mid D = 0, Z = z, A = a, X = X_i)$ for each treatment group $z \in \{0, 1\}$ and AI recommendation $a \in \{0, 1\}$, and (3) the propensity score model $e(1, X_i) := \Pr(Z = 1 \mid X = X_i)$.

Usage

```
compute_nuisance_functions_ai(
  Y,
  D,
  Z,
  A,
  V,
  d_form = D ~ .,
  y_form = Y ~ .,
  ps_form = Z ~ .,
  distribution = "bernoulli",
  n.trees = 1000,
  shrinkage = 0.01,
  interaction.depth = 1,
  ...
)
```

Arguments

`Y`	An observed outcome (binary: numeric vector of 0 or 1).
`D`	An observed decision (binary: numeric vector of 0 or 1).
`Z`	A treatment indicator (binary: numeric vector of 0 or 1).
`A`	An AI recommendation (binary: numeric vector of 0 or 1).
`V`	A `matrix` of pretreatment covariates for nuisance functions.
`d_form`	A formula for decision model where the dependent variable is `D`.
`y_form`	A formula for outcome model where the dependent variable is `Y`.
`ps_form`	A formula for propensity score model.
`distribution`	A distribution argument used in `gbm` function. Default is `"bernoulli"`.
`n.trees`	Integer specifying the total number of trees to fit used in `gbm` function.
`shrinkage`	A shrinkage parameter used in `gbm` function.
`interaction.depth`	Integer specifying the maximum depth of each tree used in `gbm` function.
`...`	Additional arguments to be passed to `gbm` function called in `crossfit`

Value

A list with the following components:

`z_models` A `data.frame` with the following columns:

- `idx` Index of observation.
- `d_pred` Predicted probability of decision.
- `y_pred` Predicted probability of outcome.
- `Z` Treatment group.
- `A` AI recommendation.

`pscore` A vector of predicted propensity scores.

Examples

```
compute_nuisance_functions_ai(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  V = NCAdata[, c("Sex", "White", "Age")],
  shrinkage = 0.01,
  n.trees = 1000
)
```

---

compute_stats *Compute Risk (Human+AI v. Human)*

Description

Compute the difference in risk between human+AI and human decision makers using difference-in-means estimators.

Usage

```
compute_stats(Y, D, Z, X = NULL, l01 = 1)
```

Arguments

`Y`	An observed outcome (binary: numeric vector of 0 or 1).
`D`	An observed decision (binary: numeric vector of 0 or 1).
`Z`	A treatment indicator (binary: numeric vector of 0 or 1).
`X`	Pretreatment covariate used for subgroup analysis (vector). Must be the same length as Y, D, Z, and A if provided. Default is NULL.
`l01`	Ratio of the loss between false positives and false negatives

Value

A tibble the following columns:

- `Z_focal`: The focal treatment indicator. '1' indicates the treatment group.
- `Z_compare`: The comparison treatment indicator. '0' indicates the control group.
- `X`: Pretreatment covariate (if provided).
- `loss_diff`: The difference in loss between human+AI and human decision
- `loss_diff_se`: The standard error of the difference in loss
- `fn_diff`: The difference in false negatives between human+AI and human decision
- `fn_diff_se`: The standard error of the difference in false negatives
- `fp_diff`: The difference in false positives between human+AI and human decision
- `fp_diff_se`: The standard error of the difference in false positives

Examples

```
compute_stats(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  X = NULL,
  l01 = 1
)
```

`compute_stats_agreement`

Agreement of Human and AI Decision Makers

Description

Estimate the impact of AI recommendations on the agreement between human decisions and AI recommendations using a difference-in-means estimator of an indicator $1\{D_i = A_i\}$.

Usage

```
compute_stats_agreement(Y, D, Z, A, X = NULL)
```

Arguments

`Y`	An observed outcome (binary: numeric vector of 0 or 1).
`D`	An observed decision (binary: numeric vector of 0 or 1).
`Z`	A treatment indicator (binary: numeric vector of 0 or 1).
`A`	An AI recommendation (binary: numeric vector of 0 or 1).
`X`	Pretreatment covariate used for subgroup analysis (vector). Must be the same length as Y, D, Z, and A if provided. Default is NULL.

Value

A tibble with the following columns:

- `X`: Pretreatment covariate (if provided).
- `agree_diff`: Difference in agreement between human decisions and AI recommendations.
- `agree_diff_se`: Standard error of the difference in agreement.

Examples

```
compute_stats_agreement(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF
)
```

`compute_stats_aipw` *Compute Risk (Human+AI v. Human)*

Description

Compute the difference in risk between human+AI and human decision makers using AIPW estimators.

Usage

```
compute_stats_aipw(Y, D, Z, nuis_funcs, true.pscore = NULL, X = NULL, l01 = 1)
```

Arguments

`Y`	An observed outcome (binary: numeric vector of 0 or 1).
`D`	An observed decision (binary: numeric vector of 0 or 1).
`Z`	A treatment indicator (binary: numeric vector of 0 or 1).
`nuis_funcs`	output from `compute_nuisance_functions`
`true.pscore`	A vector of true propensity scores (numeric), if available. Optional.
`X`	Pretreatment covariate used for subgroup analysis (vector). Must be the same length as Y, D, Z, and A if provided. Default is NULL.
`l01`	Ratio of the loss between false positives and false negatives

Value

A tibble the following columns:

- `Z_focal`: The focal treatment indicator. '1' indicates the treatment group.
- `Z_compare`: The comparison treatment indicator. '0' indicates the control group.
- `X`: Pretreatment covariate (if provided).
- `loss_diff`: The difference in loss between human+AI and human decision
- `loss_diff_se`: The standard error of the difference in loss
- `fn_diff`: The difference in false negatives between human+AI and human decision
- `fn_diff_se`: The standard error of the difference in false negatives
- `fp_diff`: The difference in false positives between human+AI and human decision
- `fp_diff_se`: The standard error of the difference in false positives

Examples

```
compute_stats_aipw(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  nuis_funcs = nuis_func,
  true.pscore = rep(0.5, nrow(NCAdata)),
  X = NULL,
  l01 = 1
)
```

---

compute_stats_subgroup

Compute Risk (Human+AI v. Human) for a Subgroup Defined by AI Recommendation

---

Description

Compute the difference in risk between human+AI and human decision makers, for a subgroup $\{A_i = a\}$, using AIPW estimators. This can be used for computing how the decision maker overrides the AI recommendation.

Usage

```
compute_stats_subgroup(
  Y,
  D,
  Z,
  A,
  a = 1,
```

```
  nuis_funcs,
  true.pscore = NULL,
  X = NULL,
  l01 = 1
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
A	An AI recommendation (binary: numeric vector of 0 or 1).
a	A specific AI recommendation value to create the subset (numeric: 0 or 1).
nuis_funcs	output from compute_nuisance_functions. If NULL, the function will compute the nuisance functions using the provided data. Note that V must be provided if nuis_funcs is NULL.
true.pscore	A vector of true propensity scores (numeric), if available. Optional.
X	Pretreatment covariate used for subgroup analysis (vector). Must be the same length as Y, D, Z, and A if provided. Default is NULL.
l01	Ratio of the loss between false positives and false negatives

Value

A tibble the following columns:

- Z_focal: The focal treatment indicator. '1' indicates the treatment group.
- Z_compare: The comparison treatment indicator. '0' indicates the control group.
- X: Pretreatment covariate (if provided).
- loss_diff: The difference in loss between human+AI and human decision
- loss_diff_se: The standard error of the difference in loss
- tn_fn_diff: The difference in true negatives and false negatives between human+AI and human decision
- tn_fn_diff_se: The standard error of the difference in true negatives and false negatives
- tp_diff: The difference in true positives between human+AI and human decision
- tp_diff_se: The standard error of the difference in true positives
- tn_diff: The difference in true negatives between human+AI and human decision
- tn_diff_se: The standard error of the difference in true negatives
- fn_diff: The difference in false negatives between human+AI and human decision
- fn_diff_se: The standard error of the difference in false negatives
- fp_diff: The difference in false positives between human+AI and human decision
- fp_diff_se: The standard error of the difference in false positives

Examples

```
compute_stats_subgroup(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  a = 1,
  nuis_funcs = nuis_func,
  true.pscore = rep(0.5, nrow(NCAdata)),
  X = NULL,
  l01 = 1
)
```

`crossfit` *Crossfitting for nuisance functions*

Description

Implement crossfitting with boosting methods and get predicted values for outcome/decision regression or propensity score models

Usage

```
crossfit(data, include_for_fit, form, ...)
```

Arguments

`data`	A `data.frame` or `matrix` to fit on.
`include_for_fit`	Boolean vector for whether or not a unit should be included in fitting (e.g. treated/control).
`form`	Formula for outcome regression/propensity score models.
`...`	Additional arguments to be passed to `gbm` function.

Value

A vector of predicted values

FTAdata *Interim Dane data with failure to appear (FTA) as an outcome*

Description

An interim dataset containing pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The data used for the paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

Usage

```
FTAdata
```

Format

A data frame with 1891 rows and 19 variables:

Z binary treatment

D ordinal decision

Y outcome

Sex male or female

White white or non-white

SexWhite the interaction between gender and race

Age age

PendingChargeAtTimeOfOffense binary variable for pending charge (felony, misdemeanor, or both) at the time of offense

NCorNonViolentMisdemeanorCharge binary variable for current non-violent felony charge

ViolentMisdemeanorCharge binary variable for current violent misdemeanor charge

ViolentFelonyCharge binary variable for current violent felony charge

NonViolentFelonyCharge binary variable for current non-violent felony charge

PriorMisdemeanorConviction binary variable for prior conviction of misdemeanor

PriorFelonyConviction binary variable for prior conviction of felony

PriorViolentConviction four-level ordinal variable for prior violent conviction

PriorSentenceToIncarceration binary variable for prior sentence to incarceration

PriorFTAInPast2Years three-level ordinal variable for FTAs from past two years

PriorFTAOlderThan2Years binary variable for FTAs from over two years ago

Staff_ReleaseRecommendation four-level ordinal variable for the DMF recommendation

g_legend *Pulling ggplot legend*

Description

Pulling ggplot legend

Usage

```
g_legend(p)
```

Arguments

p	A ggplot of which legend should be pulled.

Value

A ggplot legend.

HearingDate *Interim court event hearing date*

Description

An Interim Dane court event hearing date of Dane data in factor format. The data used for the paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

Usage

```
HearingDate
```

Format

A date variable in factor format.

hearingdate_synth *Synthetic court event hearing date*

Description

A synthetic court event hearing date

Usage

```
hearingdate_synth
```

Format

A date variable.

NCAdata *Interim Dane data with new criminal activity (NCA) as an outcome*

Description

An interim dataset containing pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The data used for the paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

Usage

```
NCAdata
```

Format

A data frame with 1891 rows and 19 variables:

Z binary treatment

D ordinal decision

Y outcome

Sex male or female

White white or non-white

SexWhite the interaction between gender and race

Age age

PendingChargeAtTimeOfOffense binary variable for pending charge (felony, misdemeanor, or both) at the time of offense

NCorNonViolentMisdemeanorCharge binary variable for current non-violent felony charge

ViolentMisdemeanorCharge binary variable for current violent misdemeanor charge

ViolentFelonyCharge binary variable for current violent felony charge

NonViolentFelonyCharge binary variable for current non-violent felony charge

PriorMisdemeanorConviction binary variable for prior conviction of misdemeanor

PriorFelonyConviction binary variable for prior conviction of felony

PriorViolentConviction four-level ordinal variable for prior violent conviction

PriorSentenceToIncarceration binary variable for prior sentence to incarceration

PriorFTAInPast2Years three-level ordinal variable for FTAs from past two years

PriorFTAOlderThan2Years binary variable for FTAs from over two years ago

Staff_ReleaseRecommendation four-level ordinal variable for the DMF recommendation

---

`NVCAdata` *Interim Dane data with new violent criminal activity (NVCA) as an outcome*

Description

An interim dataset containing pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The data used for the paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

Usage

```
NVCAdata
```

Format

A data frame with 1891 rows and 19 variables:

Z binary treatment

D ordinal decision

Y outcome

Sex male or female

White white or non-white

SexWhite the interaction between gender and race

Age age

PendingChargeAtTimeOfOffense binary variable for pending charge (felony, misdemeanor, or both) at the time of offense

NCorNonViolentMisdemeanorCharge binary variable for current non-violent felony charge

ViolentMisdemeanorCharge binary variable for current violent misdemeanor charge

ViolentFelonyCharge binary variable for current violent felony charge

NonViolentFelonyCharge binary variable for current non-violent felony charge

PriorMisdemeanorConviction binary variable for prior conviction of misdemeanor

PriorFelonyConviction binary variable for prior conviction of felony

PriorViolentConviction four-level ordinal variable for prior violent conviction

PriorSentenceToIncarceration binary variable for prior sentence to incarceration

PriorFTAInPast2Years three-level ordinal variable for FTAs from past two years

PriorFTAOlderThan2Years binary variable for FTAs from over two years ago

Staff_ReleaseRecommendation four-level ordinal variable for the DMF recommendation

---

`PlotAPCE` *Plot APCE*

Description

See Figure 4 for example.

Usage

```
PlotAPCE(
  res,
  y.max = 0.1,
  decision.labels = c("signature bond", "small cash bond", "large cash bond"),
  shape.values = c(16, 17, 15),
  col.values = c("blue", "black", "red", "brown"),
  label = TRUE,
  r.labels = c("safe", "easily\npreventable", "prevent-\nable", "risky\n"),
  label.position = c("top", "top", "top", "top"),
  top.margin = 0.01,
  bottom.margin = 0.01,
  name.group = c("Overall", "Female", "Male", "Non-white\nMale", "White\nMale"),
  label.size = 4
)
```

Arguments

`res`	A `data.frame` generated with `APCEsummary()`.
`y.max`	Maximum value of y-axis.
`decision.labels`	Labels of decisions (D).
`shape.values`	Shape of point for each decisions.
`col.values`	Color of point for each principal stratum.
`label`	A logical argument whether to specify label of each principal stratum. The default is `TRUE`.

r.labels	Label of each principal stratum.
label.position	The position of labels.
top.margin	Top margin of labels.
bottom.margin	Bottom margin of labels.
name.group	A character vector including the labels of five subgroups.
label.size	Size of label.

Value

A ggplot.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_apce <- CalAPCE(
  data = synth, mcmc.re = sample_mcmc,
  subgroup = subgroup_synth
)
sample_apce_summary <- APCEsummary(sample_apce[["APCE.mcmc"]])
PlotAPCE(sample_apce_summary,
  y.max = 0.25, decision.labels = c(
    "signature", "small cash",
    "middle cash", "large cash"
  ), shape.values = c(16, 17, 15, 18),
  col.values = c("blue", "black", "red", "brown", "purple"), label = FALSE
)
```

PlotDIMdecisions *Plot diff-in-means estimates*

Description

See Figure 2 for example.

Usage

```
PlotDIMdecisions(
  res,
  y.max = 0.2,
 decision.labels = c("signature bond  ", "small cash bond  ", "large cash bond"),
  col.values = c("grey60", "grey30", "grey6"),
  shape.values = c(16, 17, 15)
)
```

Arguments

res	A data.frame generated with CalDIMsubgroup.
y.max	Maximum value of y-axis.
decision.labels	Labels of decisions (D).
col.values	Color of point for each decisions.
shape.values	Shape of point for each decisions.

Value

A ggplot.

Examples

```
data(synth)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
res_dec <- CalDIMsubgroup(synth, subgroup = subgroup_synth)
PlotDIMdecisions(res_dec,
  decision.labels = c("signature", "small cash", "middle cash", "large cash"),
  col.values = c("grey60", "grey30", "grey6", "grey1"),
  shape.values = c(16, 17, 15, 18)
)
```

PlotDIMoutcomes *Plot diff-in-means estimates*

Description

See Figure 2 for example.

Usage

```
PlotDIMoutcomes(
  res.fta,
  res.nca,
  res.nvca,
  label.position = c("top", "top", "top"),
  top.margin = 0.01,
  bottom.margin = 0.01,
  y.max = 0.2,
  label.size = 7,
  name.group = c("Overall", "Female", "Male", "Non-white\nMale", "White\nMale")
)
```

Arguments

res.fta	A data.frame generated with CalDIMsubgroup with Y = FTA.
res.nca	A data.frame generated with CalDIMsubgroup with Y = NCA.
res.nvca	A data.frame generated with CalDIMsubgroup with Y = NVCA.
label.position	The position of labels.
top.margin	Top margin of labels.
bottom.margin	Bottom margin of labels.
y.max	Maximum value of y-axis.
label.size	Size of label.
name.group	A character vector including the labels of five subgroups.

Value

A ggplot.

Examples

```
data(synth)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
synth_fta <- synth_nca <- synth_nvca <- synth
set.seed(123)
synth_fta$Y <- sample(0:1, 1000, replace = TRUE)
synth_nca$Y <- sample(0:1, 1000, replace = TRUE)
synth_nvca$Y <- sample(0:1, 1000, replace = TRUE)
res_fta <- CalDIMsubgroup(synth_fta, subgroup = subgroup_synth)
res_nca <- CalDIMsubgroup(synth_nca, subgroup = subgroup_synth)
res_nvca <- CalDIMsubgroup(synth_nvca, subgroup = subgroup_synth)
PlotDIMoutcomes(res_fta, res_nca, res_nvca)
```

PlotFairness *Plot the principal fairness*

Description

See Figure 5 for example.

Usage

```
PlotFairness(
  res,
  top.margin = 0.01,
  y.max = 0.2,
  y.min = -0.1,
  r.labels = c("Safe", "Easily\nPreventable", "Preventable", "Risky"),
  label = TRUE
)
```

Arguments

res	The data frame generated from CalFairness.
top.margin	The index of subgroups (within the output of CalAPCE/CalAPCEparallel) that corresponds to the protected attributes.
y.max	Maximum value of y-axis.
y.min	Minimum value of y-axis.
r.labels	Label of each principal stratum.
label	A logical argument whether to specify label.

Value

A data.frame of the delta.

Examples

```
data(synth)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
sample_apce <- CalAPCE(
  data = synth, mcmc.re = sample_mcmc, subgroup = subgroup_synth,
  burnin = 0
)
sample_fair <- CalFairness(sample_apce)
PlotFairness(sample_fair, y.max = 0.4, y.min = -0.4, r.labels = c(
  "Safe", "Preventable 1",
  "Preventable 2", "Preventable 3", "Risky"
))
```

PlotOptimalDecision *Plot optimal decision*

Description

See Figure 6 for example.

Usage

```
PlotOptimalDecision(res, colname.d, idx = NULL)
```

Arguments

res	The data frame generated from CalOptimalDecision.
colname.d	The column name of decision to be plotted.
idx	The row index of observations to be included. The default is all the observations from the data.

Value

A ggplot.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
sample_optd <- CalOptimalDecision(
  data = synth, mcmc.re = sample_mcmc,
  c0.ls = seq(0, 5, 1), c1.ls = seq(0, 5, 1),
  size = 1
) # adjust the size
sample_optd$cash <- sample_optd$d1 + sample_optd$d2 + sample_optd$d3
PlotOptimalDecision(sample_optd, "cash")
```

PlotPS *Plot the proportion of principal strata (R)*

Description

See Figure 3 for example.

Usage

```
PlotPS(
  res,
  y.min = 0,
  y.max = 0.75,
  col.values = c("blue", "black", "red", "brown"),
  label = TRUE,
  r.labels = c("safe", " easily             \n preventable    ",
    "\n          preventable\n", "  risky"),
  label.position = c("top", "top", "top", "bottom"),
  top.margin = 0.02,
  bottom.margin = 0.02,
  label.size = 6.5
)
```

Arguments

res	A data.frame generated with CalPS.
y.min	Minimum value of y-axis.
y.max	Maximum value of y-axis.
col.values	Color of point for each principal stratum.
label	A logical argument whether to specify label of each principal stratum. The default is TRUE.
r.labels	Label of each principal stratum.
label.position	The position of labels.
top.margin	Top margin of labels.
bottom.margin	Bottom margin of labels.
label.size	Size of label.

Value

A ggplot.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
subgroup_synth <- list(
  1:nrow(synth), which(synth$Sex == 0), which(synth$Sex == 1),
  which(synth$Sex == 1 & synth$White == 0), which(synth$Sex == 1 & synth$White == 1)
)
sample_apce <- CalAPCE(
  data = synth, mcmc.re = sample_mcmc,
  subgroup = subgroup_synth
)
sample_ps <- CalPS(sample_apce[["P.R.mcmc"]])
PlotPS(sample_ps, col.values = c("blue", "black", "red", "brown", "purple"), label = FALSE)
```

PlotSpilloverCRT *Plot conditional randomization test*

Description

See Figure S8 for example.

Usage

```
PlotSpilloverCRT(res)
```

Arguments

res	A list generated with SpilloverCRT.

Value

A ggplot

Examples

```
data(synth)
data(hearingdate_synth)
crt <- SpilloverCRT(D = synth$D, Z = synth$Z, CourtEvent_HearingDate = hearingdate_synth)
PlotSpilloverCRT(crt)
```

PlotSpilloverCRTpower *Plot power analysis of conditional randomization test*

Description

See Figure S9 for example.

Usage

```
PlotSpilloverCRTpower(res)
```

Arguments

res	A data.frame generated with SpilloverCRTpower.

Value

A ggplot

Examples

```
data(synth)
data(hearingdate_synth)
crt_power <- SpilloverCRTpower(
  D = synth$D, Z = synth$Z,
  CourtEvent_HearingDate = hearingdate_synth,
  size = 1
) # adjust the size
PlotSpilloverCRTpower(crt_power)
```

PlotStackedBar *Stacked barplot for the distribution of the decision given psa*

Description

See Figure 1 for example.

Usage

```
PlotStackedBar(
  data,
  fta.label = "FTAScore",
  nca.label = "NCAScore",
  nvca.label = "NVCAFlag",
  d.colors = c("grey60", "grey30", "grey10"),
  d.labels = c("signature bond", "small cash bond", "large cash bond"),
  legend.position = "none"
)
```

Arguments

data	A data.frame of which columns includes an ordinal decision (D), and psa variables (fta, nca, and nvca).
fta.label	Column name of fta score in the data. The default is "FTAScore".
nca.label	Column name of nca score in the data. The default is "NCAScore".
nvca.label	Column name of nvca score in the data. The default is "NVCAFlag".
d.colors	The color of each decision.
d.labels	The label of each decision.
legend.position	The position of legend. The default is "none".

Value

A list of three ggplots.

Examples

```
data(psa_synth)
# Control group (PSA not provided)
PlotStackedBar(psa_synth[psa_synth$Z == 0, ], d.colors = c(
  "grey80", "grey60",
  "grey30", "grey10"
), d.labels = c(
  "signature", "small",
  "middle", "large"
))
# Treated group (PSA provided)
PlotStackedBar(psa_synth[psa_synth$Z == 0, ], d.colors = c(
  "grey80", "grey60",
  "grey30", "grey10"
), d.labels = c(
  "signature", "small",
  "middle", "large"
))
```

PlotStackedBarDMF *Stacked barplot for the distribution of the decision given DMF recommendation*

Description

See Figure 1 for example.

Usage

```
PlotStackedBarDMF(
  data,
  dmf.label = "dmf",
  dmf.category = NULL,
  d.colors = c("grey60", "grey30", "grey10"),
  d.labels = c("signature bond", "small cash bond", "large cash bond"),
  legend.position = "none"
)
```

Arguments

data	A data.frame of which columns includes a binary treatment (Z; PSA provision), an ordinal decision (D), and DMF recommendation.
dmf.label	Column name of DMF recommendation in the data. The default is "dmf".

dmf.category	The name of each category of DMF recommendation.
d.colors	The color of each decision.
d.labels	The label of each decision.
legend.position	The position of legend. The default is "none".

Value

A list of three ggplots.

Examples

```
data(psa_synth)
PlotStackedBarDMF(psa_synth, dmf.label = "DMF", d.colors = c(
  "grey80",
  "grey60", "grey30", "grey10"
), d.labels = c(
  "signature",
  "small", "middle", "large"
))
```

PlotUtilityDiff *Plot utility difference*

Description

See Figure 7 for example.

Usage

```
PlotUtilityDiff(res, idx = NULL)
```

Arguments

res	The data frame generated from CalUtilityDiff.
idx	The row index of observations to be included. The default is all the observations from the data.

Value

A ggplot.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
synth_dmf <- sample(0:1, nrow(synth), replace = TRUE) # random dmf recommendation
sample_utility <- CalOptimalDecision(
  data = synth, mcmc.re = sample_mcmc,
  c0.ls = seq(0, 5, 1), c1.ls = seq(0, 5, 1),
  dmf = synth_dmf, size = 1
) # adjust the size
PlotUtilityDiff(sample_utility)
```

PlotUtilityDiffCI *Plot utility difference with 95% confidence interval*

Description

See Figure S17 for example.

Usage

```
PlotUtilityDiffCI(res)
```

Arguments

res	The second data frame (res.mcmc) generated from `CalUtilityDiff(include.utility.diff.mcmc = TRUE)`.

Value

A ggplot.

Examples

```
data(synth)
sample_mcmc <- AiEvalmcmc(data = synth, n.mcmc = 10)
synth_dmf <- sample(0:1, nrow(synth), replace = TRUE) # random dmf recommendation
sample_utility <- CalOptimalDecision(
  data = synth, mcmc.re = sample_mcmc,
  c0.ls = seq(0, 5, 1), c1.ls = seq(0, 5, 1),
  dmf = synth_dmf, size = 1, # adjust the size
  include.utility.diff.mcmc = TRUE
)
PlotUtilityDiffCI(sample_utility$res.mcmc)
```

plot_agreement *Visualize Agreement*

Description

Visualize the agreement between human decisions and AI recommendations using a difference-in-means estimator of an indicator $1\{D_i = A_i\}$. Generate a plot based on the overall agreement and subgroup-specific agreement.

Usage

```
plot_agreement(
  Y,
  D,
  Z,
  A,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2",
  x.order = NULL,
  p.title = NULL,
  p.lb = -0.3,
  p.ub = 0.3,
  y.lab = "Impact of PSA"
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
A	An AI recommendation (binary: numeric vector of 0 or 1).
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".
label.subgroup2	A label for subgroup2 (character). Default "Subgroup 2".
x.order	An order for the x-axis (character vector). Default NULL.
p.title	A title for the plot (character). Default NULL.
p.lb	A lower bound for the y-axis (numeric). Default -0.3.
p.ub	An upper bound for the y-axis (numeric). Default 0.3.
y.lab	A label for the y-axis (character). Default "Impact of PSA".

Value

A ggplot object.

Examples

```
plot_agreement(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender",
  x.order = c("Overall", "Non-white", "White", "Female", "Male")
)
```

---

plot_diff_ai_aipw *Visualize Difference in Risk (AI v. Human)*

Description

Visualize the difference in risk between AI and human decision makers using AIPW estimators. Generate a plot based on the overall and subgroup-specific results.

Usage

```
plot_diff_ai_aipw(
  Y,
  D,
  Z,
  V = NULL,
  A,
  z_compare = 0,
  l01 = 1,
  nuis_funcs = NULL,
  nuis_funcs_ai = NULL,
  true.pscore = NULL,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2",
  x.order = NULL,
  zero.line = TRUE,
  arrows = TRUE,
  y.min = -Inf,
  p.title = NULL,
```

```
  p.lb = -1,
  p.ub = 1,
  y.lab = "PSA versus Human",
  p.label = c("PSA worse", "PSA better")
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
V	A matrix of pretreatment covariates (numeric matrix). Optional.
A	An observed AI recommendation (binary: numeric vector of 0 or 1).
z_compare	A compare treatment indicator (numeric). Default 0.
l01	Ratio of the loss between false positives and false negatives. Default 1.
nuis_funcs	output from compute_nuisance_functions. If NULL, the function will compute the nuisance functions using the provided data. Note that V must be provided if nuis_funcs is NULL.
nuis_funcs_ai	output from compute_nuisance_functions_ai
true.pscore	A vector of true propensity scores (numeric), if available. Optional.
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".
label.subgroup2	A label for subgroup2 (character). Default "Subgroup 2".
x.order	An order for the x-axis (character vector). Default NULL.
zero.line	A logical indicating whether to include a zero line. Default TRUE.
arrows	A logical indicating whether to include arrows. Default TRUE.
y.min	A lower bound for the y-axis (numeric). Default -Inf.
p.title	A title for the plot (character). Default NULL.
p.lb	A lower bound for the y-axis (numeric). Default -0.2.
p.ub	An upper bound for the y-axis (numeric). Default 0.2.
y.lab	A label for the y-axis (character). Default "PSA versus Human".
p.label	A vector of two labels for the annotations (character). Default c("PSA harms", "PSA helps").

Value

A ggplot object.

Examples

```
plot_diff_ai_aipw(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  z_compare = 0,
  nuis_funcs = nuis_func,
  nuis_funcs_ai = nuis_func_ai,
  true.pscore = rep(0.5, nrow(NCAdata)),
  l01 = 1,
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender",
  x.order = c("Overall", "Non-white", "White", "Female", "Male"),
  zero.line = TRUE, arrows = TRUE, y.min = -Inf,
  p.title = NULL, p.lb = -0.3, p.ub = 0.3
)
```

---

plot_diff_human *Visualize Difference in Risk (Human+AI v. Human)*

Description

Visualize the the difference in risk between human+AI and human decision makers using difference-in-means estimators. Generate a plot based on the overall agreement and subgroup-specific agreement.

Usage

```
plot_diff_human(
  Y,
  D,
  Z,
  l01 = 1,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2",
  x.order = NULL,
  p.title = NULL,
  p.lb = -1,
  p.ub = 1,
  y.lab = "Impact of PSA",
  p.label = c("PSA harms", "PSA helps")
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
l01	Ratio of the loss between false positives and false negatives. Default 1.
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".
label.subgroup2	A label for subgroup2 (character). Default "Subgroup 2".
x.order	An order for the x-axis (character vector). Default NULL.
p.title	A title for the plot (character). Default NULL.
p.lb	A lower bound for the y-axis (numeric). Default -1.
p.ub	An upper bound for the y-axis (numeric). Default 1.
y.lab	A label for the y-axis (character). Default "Impact of PSA".
p.label	A vector of two labels for the annotations (character). Default c("PSA harms", "PSA helps").

Value

A ggplot object.

Examples

```
plot_diff_human(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  l01 = 1,
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender",
  x.order = c("Overall", "Non-white", "White", "Female", "Male"),
  p.title = NULL, p.lb = -0.3, p.ub = 0.3
)
```

plot_diff_human_aipw *Visualize Difference in Risk (Human+AI v. Human)*

Description

Visualize the difference in risk between human+AI and human decision makers using AIPW estimators. Generate a plot based on the overall agreement and subgroup-specific agreement.

Usage

```
plot_diff_human_aipw(
  Y,
  D,
  Z,
  V = NULL,
  l01 = 1,
  nuis_funcs = NULL,
  true.pscore = NULL,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2",
  x.order = NULL,
  p.title = NULL,
  p.lb = -1,
  p.ub = 1,
  y.lab = "Impact of PSA",
  p.label = c("PSA harms", "PSA helps")
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
V	A matrix of pretreatment covariates (numeric matrix). Optional.
l01	Ratio of the loss between false positives and false negatives. Default 1.
nuis_funcs	output from compute_nuisance_functions. If NULL, the function will compute the nuisance functions using the provided data. Note that V must be provided if nuis_funcs is NULL.
true.pscore	A vector of true propensity scores (numeric), if available. Optional.
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".

label.subgroup2
: A label for subgroup2 (character). Default "Subgroup 2".

x.order
: An order for the x-axis (character vector). Default NULL.

p.title
: A title for the plot (character). Default NULL.

p.lb
: A lower bound for the y-axis (numeric). Default -1.

p.ub
: An upper bound for the y-axis (numeric). Default 1.

y.lab
: A label for the y-axis (character). Default "Impact of PSA".

p.label
: A vector of two labels for the annotations (character). Default c("PSA harms", "PSA helps").

Value

A ggplot object.

Examples

```
plot_diff_human_aipw(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  nuis_funcs = nuis_func,
  true.pscore = rep(0.5, nrow(NCAdata)),
  l01 = 1,
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender",
  x.order = c("Overall", "Non-white", "White", "Female", "Male"),
  p.title = NULL, p.lb = -0.3, p.ub = 0.3
)
```

---

plot_diff_subgroup *Visualize Difference in Risk (Human+AI v. Human) for a Subgroup Defined by AI Recommendation*

Description

Visualize the the difference in risk between human+AI and human decision makers using AIPW estimators, for a subgroup defined by AI recommendation. Generate a plot based on the overall agreement and subgroup-specific agreement.

Usage

```
plot_diff_subgroup(
  Y,
  D,
  Z,
  A,
  a = 1,
  V = NULL,
  l01 = l01,
  nuis_funcs = NULL,
  true.pscore = NULL,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2",
  x.order = NULL,
  p.title = NULL,
  p.lb = -1,
  p.ub = 1,
  y.lab = "Impact of PSA",
  p.label = c("Human correct", "PSA correct"),
  label = "TNP - FNP",
  metrics = c("Misclassification Rate", "False Negative Proportion",
    "False Positive Proportion")
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
A	An AI recommendation (binary: numeric vector of 0 or 1).
a	A specific AI recommendation value to create the subset (numeric: 0 or 1).
V	A matrix of pretreatment covariates (numeric matrix). Optional.
l01	Ratio of the loss between false positives and false negatives. Default 1.
nuis_funcs	output from compute_nuisance_functions. If NULL, the function will compute the nuisance functions using the provided data. Note that V must be provided if nuis_funcs is NULL.
true.pscore	A vector of true propensity scores (numeric), if available. Optional.
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".
label.subgroup2	A label for subgroup2 (character). Default "Subgroup 2".

x.order	An order for the x-axis (character vector). Default NULL.
p.title	A title for the plot (character). Default NULL.
p.lb	A lower bound for the y-axis (numeric). Default -1.
p.ub	An upper bound for the y-axis (numeric). Default 1.
y.lab	A label for the y-axis (character). Default "Impact of PSA".
p.label	A vector of two labels for the annotations (character). Default c("Human correct", "PSA correct").
label	A label for the plot (character). Default "TNP - FNP".
metrics	A vector of metrics to include in the plot (character). Default c("Misclassification Rate", "False Negative Proportion", "False Positive Proportion").

Value

A ggplot object.

Examples

```
plot_diff_subgroup(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  a = 1,
  l01 = 1,
  nuis_funcs = nuis_func,
  true.pscore = rep(0.5, nrow(NCAdata)),
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender",
  x.order = c("Overall", "Non-white", "White", "Female", "Male"),
  p.title = NULL, p.lb = -0.5, p.ub = 0.5,
  label = "TNP - FNP",
 metrics = c("True Negative Proportion (TNP)", "False Negative Proportion (FNP)", "TNP - FNP")
)
```

---

plot_preference *Visualize Preference*

Description

Compute the difference in risk between AI and human decision makers using AIPW estimators over a set of loss ratios, and then visualize when we prefer human over AI decision makers. Generate a plot based on the overall and subgroup-specific results.

Usage

```
plot_preference(
  Y,
  D,
  Z,
  V = NULL,
  A,
  z_compare = 0,
  true.pscore = NULL,
  nuis_funcs = NULL,
  nuis_funcs_ai = NULL,
  l01_seq = 10^seq(-2, 2, length.out = 100),
  alpha = 0.05,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2",
  x.order = NULL,
  p.title = NULL,
  legend.position = "none",
  p.label = c("AI-alone preferred", "Human-alone preferred", "Ambiguous")
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
V	A matrix of pretreatment covariates (numeric matrix). Optional.
A	An observed AI recommendation (binary: numeric vector of 0 or 1).
z_compare	A compare treatment indicator (numeric). Default 0.
true.pscore	A vector of true propensity scores (numeric), if available. Optional.
nuis_funcs	output from compute_nuisance_functions. If NULL, the function will compute the nuisance functions using the provided data. Note that V must be provided if nuis_funcs is NULL.
nuis_funcs_ai	output from compute_nuisance_functions_ai
l01_seq	A candidate list of ratio of the loss between false positives and false negatives. Default 10^seq(-2, 2, length.out = 100).
alpha	A significance level (numeric). Default 0.05.
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".
label.subgroup2	A label for subgroup2 (character). Default "Subgroup 2".

x.order	An order for the x-axis (character vector). Default NULL.
p.title	A title for the plot (character). Default NULL.
legend.position	Position of the legend (character).
p.label	A vector of three labels for the annotations (character). Default c("AI-alone preferred", "Human-alone preferred", "Ambiguous").

Value

A ggplot object.

Examples

```
plot_preference(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  z_compare = 0,
  nuis_funcs = nuis_func,
  nuis_funcs_ai = nuis_func_ai,
  true.pscore = rep(0.5, nrow(NCAdata)),
  l01_seq = 10^seq(-2, 2, length.out = 10),
  alpha = 0.05,
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender",
  x.order = c("Overall", "Non-white", "White", "Female", "Male"),
  p.title = NULL, legend.position = "none",
  p.label = c("AI-alone preferred", "Human-alone preferred", "Ambiguous")
)
```

PSAdata *Interim Dane PSA data*

Description

An interim dataset containing a binary treatment (Z), ordinal decision (D), three PSA variables (FTAScore, NCAScore, and NVCAFlag), DMF recommendation, and two pre-treatment covariates (binary indicator for gender; binary indicator for race). The data used for the paper, and made available here, are interim, based on only half of the observations in the study and (for those observations) only half of the study follow-up period. We use them only to illustrate methods, not to draw substantive conclusions.

Usage

```
PSAdata
```

Format

A data frame with 1891 rows and 7 variables:

Z binary treatment

D ordinal decision

FTAScore FTA score

NCAScore NCA score

NVCAFlag NVCA flag

DMF DMF recommendation

Sex male or female

White white or non-white

`psa_synth` *Synthetic PSA data*

Description

A synthetic dataset containing a binary treatment (Z), ordinal decision (D), three PSA variables (FTAScore, NCAScore, and NVCAFlag), and DMF recommendation.

Usage

```
psa_synth
```

Format

A data frame with 1000 rows and 4 variables:

Z binary treatment

D ordinal decision

FTAScore FTA score

NCAScore NCA score

NVCAFlag NVCA flag

DMF DMF recommendation

SpilloverCRT *Conduct conditional randomization test*

Description

See S3.1 for more details.

Usage

```
SpilloverCRT(D, Z, CourtEvent_HearingDate, n = 100, seed.number = 12345)
```

Arguments

D	A numeric vector of judge's decision.
Z	A numeric vector of treatment variable.
CourtEvent_HearingDate	The court event hearing date.
n	Number of permutations.
seed.number	An integer for random number generator.

Value

A `list` of the observed and permuted test statistics and its p-value.

Examples

```
data(synth)
data(hearingdate_synth)
crt <- SpilloverCRT(D = synth$D, Z = synth$Z, CourtEvent_HearingDate = hearingdate_synth)
```

SpilloverCRTpower *Conduct power analysis of conditional randomization test*

Description

See S3.2 for more details.

Usage

```
SpilloverCRTpower(
  D,
  Z,
  CourtEvent_HearingDate,
  n = 4,
  m = 4,
  size = 2,
  cand_omegaZtilde = seq(-1.5, 1.5, by = 0.5)
)
```

Arguments

D	A numeric vector of judge's decision.
Z	A numeric vector of treatment variable.
CourtEvent_HearingDate	The court event hearing date.
n	Number of permutations.
m	Number of permutations.
size	The number of parallel computing. The default is 2.
cand_omegaZtilde	Candidate values

Value

A data.frame of the result of power analysis.

Examples

```
data(synth)
data(hearingdate_synth)
crt_power <- SpilloverCRTpower(
  D = synth$D, Z = synth$Z,
  CourtEvent_HearingDate = hearingdate_synth,
  size = 1
) # adjust the size
```

---

synth *Synthetic data*

Description

A synthetic dataset containing pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y).

Usage

```
synth
```

Format

A data frame with 1000 rows and 11 variables:

Z binary treatment

D ordinal decision

Y outcome

Sex male or female

White white or non-white

Age age

CurrentViolentOffense binary variable for current violent offense

PendingChargeAtTimeOfOffense binary variable for pending charge (felony, misdemeanor, or both) at the time of offense

PriorMisdemeanorConviction binary variable for prior conviction of misdemeanor

PriorFelonyConviction binary variable for prior conviction of felony

PriorViolentConviction four-level ordinal variable for prior violent conviction

---

table_agreement *Table of Agreement*

Description

Estimate the impact of AI recommendations on the agreement between human decisions and AI recommendations using a difference-in-means estimator of an indicator $1\{D_i = A_i\}$. Generate a table based on the overall agreement and subgroup-specific agreement.

Usage

```
table_agreement(
  Y,
  D,
  Z,
  A,
  subgroup1,
  subgroup2,
  label.subgroup1 = "Subgroup 1",
  label.subgroup2 = "Subgroup 2"
)
```

Arguments

Y	An observed outcome (binary: numeric vector of 0 or 1).
D	An observed decision (binary: numeric vector of 0 or 1).
Z	A treatment indicator (binary: numeric vector of 0 or 1).
A	An AI recommendation (binary: numeric vector of 0 or 1).
subgroup1	A pretreatment covariate used for subgroup analysis (vector).
subgroup2	A pretreatment covariate used for subgroup analysis (vector).
label.subgroup1	A label for subgroup1 (character). Default "Subgroup 1".
label.subgroup2	A label for subgroup2 (character). Default "Subgroup 2".

Value

A tibble with the following columns:

- cov: Subgroup label.
- X: Subgroup value.
- agree_diff: Difference in agreement between human decisions and AI recommendations.
- agree_diff_se: Standard error of the difference in agreement.
- ci_lb: Lower bound of the 95% confidence interval.
- ci_ub: Upper bound of the 95% confidence interval.

Examples

```
table_agreement(
  Y = NCAdata$Y,
  D = ifelse(NCAdata$D == 0, 0, 1),
  Z = NCAdata$Z,
  A = PSAdata$DMF,
  subgroup1 = ifelse(NCAdata$White == 1, "White", "Non-white"),
  subgroup2 = ifelse(NCAdata$Sex == 1, "Male", "Female"),
  label.subgroup1 = "Race",
  label.subgroup2 = "Gender"
)
```

TestMonotonicity *Test monotonicity*

Description

Test monotonicity using frequentist analysis

Usage

```
TestMonotonicity(data)
```

Arguments

data	A data.frame or matrix of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.

Value

Message indicating whether the monotonicity assumption holds.

Examples

```
data(synth)
TestMonotonicity(synth)
```

TestMonotonicityRE *Test monotonicity with random effects*

Description

Test monotonicity using frequentist analysis with random effects for the hearing date of the case.

Usage

```
TestMonotonicityRE(data, fixed, random)
```

Arguments

data	A data.frame or matrix of which columns consists of pre-treatment covariates, a binary treatment (Z), an ordinal decision (D), and an outcome variable (Y). The column names of the latter three should be specified as "Z", "D", and "Y" respectively.
fixed	A formula for the fixed-effects part of the model to fit.
random	A formula for the random-effects part of the model to fit.

Value

Message indicating whether the monotonicity assumption holds.

References

Imai, K., Jiang, Z., Greiner, D.J., Halen, R., and Shin, S. (2023). "Experimental evaluation of algorithm-assisted human decision-making: application to pretrial public safety assessment." Journal of the Royal Statistical Society: Series A. <DOI:10.1093/jrsssa/qnad010>.

Examples

```
data(synth)
data(hearingdate_synth)
synth$CourtEvent_HearingDate <- hearingdate_synth
TestMonotonicityRE(synth,
  fixed = "Y ~ Sex + White + Age +
                  CurrentViolentOffense + PendingChargeAtTimeOfOffense +
                  PriorMisdemeanorConviction + PriorFelonyConviction +
                  PriorViolentConviction + D",
  random = "~ 1|CourtEvent_HearingDate"
)
```

Index