Package 'cWise'

September 28, 2026

Type Package

Title Crosswise Models for Sensitive Survey Questions

Version 0.1.0

Description Implements a bias-corrected crosswise estimator and its extensions
for sensitive survey questions. The methods are described in
Atsusaka and Stevenson (2023). ``A bias-corrected estimator for the crosswise model with inattentive respondents" <doi:10.1017/pan.2021.43>.

License GPL-3

URL https://github.com/YukiAtsusaka/cWise,
https://www.atsusaka.org/cwise/

BugReports https://github.com/YukiAtsusaka/cWise/issues

Depends R (>= 3.5.0)

Imports dplyr, ggplot2, mvtnorm (>= 1.1-1), scales

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.3.3

NeedsCompilation no

Author Yuki Atsusaka [aut, cre],
Kolbe Dumas [aut]

Maintainer Yuki Atsusaka <atsusaka@uh.edu>

Repository CRAN

Date/Publication 2026-09-28 09:00:24 UTC

Contents

`bc_est` *bc_est*

Description

`bc_est` is used to apply a bias-corrected crosswise estimator to survey data.

Usage

```
bc_est(Y, A, p, p.prime, weight, data, seed = NULL)

bc.est(...)
```

Arguments

`Y`	a vector of binary responses in the crosswise question (Y=1 if TRUE-TRUE or FALSE-FALSE, Y=0 otherwise).
`A`	a vector of binary responses in the anchor question (A=1 if TRUE-TRUE or FALSE-FALSE, A=0 otherwise).
`p`	an auxiliary probability for the crosswise question.
`p.prime`	an auxiliary probability for the anchor question.
`weight`	an optional vector specifying sample weights in data.
`data`	a data frame containing information from the crosswise model (Y, A, weight).
`seed`	Optional integer used to reproduce the bootstrap uncertainty estimate. When supplied, the caller's random-number state is restored after the function returns.
`...`	Arguments passed to `bc_est()`.

Value

A list with:

Results A two-row matrix of naive and bias-corrected prevalence estimates, standard errors, and 95 percent confidence intervals.

Stats A matrix containing the estimated attentive-response rate and the number of complete observations.

References

Atsusaka, Y. and Stevenson, R. T. (2023). The crosswise model for sensitive survey questions. doi:10.1017/pan.2021.43.

See Also

[cmBound()] for a sensitivity analysis of naive crosswise estimates.

Examples

```
bc_est(Y=Y, A=A, p=0.15, p.prime=0.15, data=cmdata)
bc_est(Y=Y, A=A, weight=weight, p=0.15, p.prime=0.15, data=cmdata)
```

`cmBound` *cmBound*

Description

`cmBound` is used to apply a sensitivity analysis and visualize the sensitivity bounds for naive crosswise estimates. The sensitivity analysis assumes that inattentive respondents choose either response option with equal probability.

Usage

```
cmBound(lambda.hat, p, N, dq = NULL, N.dq = NULL)
```

Arguments

`lambda.hat`	a value for the observed crosswise proportion: Prop(TRUE-TRUE or FALSE-FALSE).
`p`	an auxiliary probability for the crosswise question. It must be a single finite value strictly between 0 and 1, excluding 0.5.
`N`	a value for the number of survey respondents in the crosswise (and anchor) question.
`dq`	a value for a point estimate from direct questioning (if available).
`N.dq`	a value for the number of survey respondents in direct questioning (if available).

Value

A ggplot object visualizing the result of sensitivity analysis.

References

Atsusaka, Y. and Stevenson, R. T. (2023). The crosswise model for sensitive survey questions. doi:10.1017/pan.2021.43.

See Also

[bc_est()] for bias-corrected prevalence estimation.

Examples

```
p <- cmBound(lambda.hat=0.6385, p=0.25, N=310, dq=0.073, N.dq=310)
p

p <- p + ggplot2::ggtitle("Sensitivity Analysis") +
         ggplot2::theme(plot.title = ggplot2::element_text(size=20, face="bold"))
```

---

cmdata *Simulated Crosswise Survey Data I*

Description

Typical, artificially generated toy data set that comes with the binary response in the crosswise question (Y=1 if TRUE-TRUE or FALSE-FALSE, Y=0 otherwise), the binary response in the anchor question, sample weights, and auxiliary probabilities in the crosswise and anchor questions, respectively (these columns are only included here for an illustrative purpose).

Usage

```
data(cmdata)
```

Format

an object of class `"data.frame"`

Y a binary indicator for the crosswise item (TRUE-TRUE or FALSE-FALSE) in the crosswise question

A a binary indicator for the crosswise item (TRUE-TRUE or FALSE-FALSE) in the anchor question

p an auxiliary probability in the crosswise question

p.prime an auxiliary probability in the anchor question

References

This data set was artificially created for the cWise package.

Examples

```
data(cmdata)
head(cmdata)
```

`cmdata2` *Simulated Crosswise Survey Data II*

Description

Typical, artificially generated toy data set that comes with the binary response in the crosswise question (Y=1 if TRUE-TRUE or FALSE-FALSE, Y=0 otherwise), the binary response in the anchor question, two covariates, and auxiliary probabilities in the crosswise and anchor questions, respectively (these columns are only included here for an illustrative purpose).

Usage

```
data(cmdata2)
```

Format

an object of class `"data.frame"`

Y a binary indicator for the crosswise item (TRUE-TRUE or FALSE-FALSE) in the crosswise question

female A binary covariate representing being a female

age A discrete covariate representing age

A a binary indicator for the crosswise item (TRUE-TRUE or FALSE-FALSE) in the anchor question

p an auxiliary probability in the crosswise question

p.prime an auxiliary probability in the anchor question

References

This data set was artificially created for the cWise package.

Examples

```
data(cmdata2)
head(cmdata2)
```

`cmdata3` *Simulated Crosswise Survey Data III*

Description

Typical, artificially generated toy data set that comes with the binary response in the crosswise question (Y=1 if TRUE-TRUE or FALSE-FALSE, Y=0 otherwise), the binary response in the anchor question, the outcome variable, two covariates, and auxiliary probabilities in the crosswise and anchor questions, respectively (these columns are only included here for an illustrative purpose).

Usage

```
data(cmdata3)
```

Format

an object of class `"data.frame"`

V an outcome variable

female A binary covariate representing being a female

age A discrete covariate representing age

Y a binary indicator for the crosswise item (TRUE-TRUE or FALSE-FALSE) in the crosswise question

A a binary indicator for the crosswise item (TRUE-TRUE or FALSE-FALSE) in the anchor question

p an auxiliary probability in the crosswise question

p.prime an auxiliary probability in the anchor question

References

This data set was artificially created for the cWise package.

Examples

```
data(cmdata3)
head(cmdata3)
```

cmpredict *cmpredict*

Description

Perform a post-estimation prediction with uncertainty quantification via parametric bootstrap

Usage

```
cmpredict(
  out,
  newdata = NULL,
  zval = NULL,
  typical = NULL,
  nsim = 1000L,
  seed = NULL,
  draws = FALSE
)
```

Arguments

out	An output of cmreg.
newdata	A named data frame, list, or vector supplying values for every covariate in the fitted formula. A data frame returns one prediction row per row of ‘newdata‘.
zval	Optional named vector for one covariate to vary. Its name must match a formula variable; all remaining covariates are supplied through ‘newdata‘ or ‘typical‘.
typical	Optional named vector or list of fixed covariate values. This is a convenience alternative to ‘newdata‘ when used with ‘zval‘.
nsim	Number of parametric-bootstrap draws.
seed	Optional integer seed for reproducible bootstrap draws. When set, the caller’s RNG state is restored before returning.
draws	If ‘TRUE‘, attach the raw bootstrap draw matrix as a ‘"draws"‘ attribute on the returned data frame.

Value

A data frame with ‘estimate‘, ‘conf.low‘, and ‘conf.high‘ columns, with one row for each prediction scenario. When ‘draws = TRUE‘, its ‘"draws"‘ attribute contains the raw parametric-bootstrap matrix.

References

Atsusaka, Y. and Stevenson, R. T. (2023). The crosswise model for sensitive survey questions. doi:10.1017/pan.2021.43.

See Also

[cmreg()] to fit the required outcome-model object and [cmpredict_p()] for predictions from a predictor model.

Examples

```
m <- cmreg(Y ~ female + age, anchor = A, p = 0.1, p.prime = 0.15,
           data = cmdata2)
predictions <- cmpredict(m, typical = c(age = 30),
                         zval = c(female = 0, female = 1))
predictions
```

`cmpredict_p` *cmpredict_p*

Description

Perform a post-estimation prediction with uncertainty quantification via parametric bootstrap

Usage

```
cmpredict_p(
  out,
  newdata = NULL,
  zval = NULL,
  typical = NULL,
  type = c("response", "link"),
  nsim = 1000L,
  seed = NULL,
  draws = FALSE
)

cmpredict.p(...)
```

Arguments

`out`	An output of `cmreg_p`.
`newdata`	A named data frame, list, or vector supplying values for every covariate in the fitted formula. A data frame returns absent/present latent- trait scenarios for every row.
`zval`	Optional named vector for one covariate to vary. Its name must match a formula variable; all remaining covariates are supplied through ‘newdata‘ or ‘typical‘.
`typical`	Optional named vector or list of fixed covariate values. This is a convenience alternative to ‘newdata‘ when used with ‘zval‘.
`type`	Prediction scale: ‘"response"‘ (default) or ‘"link"‘. The Gaussian outcome model has an identity link, so these are currently equal.

nsim	Number of parametric-bootstrap draws.
seed	Optional integer seed for reproducible bootstrap draws. When set, the caller's RNG state is restored before returning.
draws	If 'TRUE', attach the raw bootstrap draw matrix as a '"draws"' attribute on the returned data frame.
...	Arguments passed to cmpredict_p().

Value

A data frame with 'estimate', 'conf.low', and 'conf.high' columns. Each input scenario contributes an absent-trait row followed by a present- trait row. When 'draws = TRUE', its '"draws"' attribute contains the raw parametric-bootstrap matrix.

References

Atsusaka, Y. and Stevenson, R. T. (2023). The crosswise model for sensitive survey questions. doi:10.1017/pan.2021.43.

See Also

[cmreg_p()] to fit the required predictor-model object and [cmpredict()] for predictions from an outcome model.

Examples

```
# Keep the fitted example small: this likelihood is deliberately
# computationally intensive and package examples run during CRAN checks.
example_data <- cmdata3[seq_len(30L), ]
m2 <- cmreg_p(V ~ age + female, crosswise = Y, anchor = A, p = 0.1,
              p.prime = 0.15, data = example_data, n.start = 1L)
predictions <- cmpredict_p(
  m2, newdata = data.frame(age = 30, female = 1), nsim = 20L, seed = 1L
)
predictions
```

---

cmreg *cmreg*

Description

cmreg is used to run a regression with the latent sensitive trait as an outcome.

Usage

```
cmreg(
  formula,
  anchor,
  p,
  p.prime,
  data,
  start = NULL,
  n.start = 3L,
  control = list()
)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted. Ex. Crosswise response ~ Covariates. The anchor response is supplied separately through ‘anchor‘.
anchor	Unquoted column name (or a single character column name) for the anchor response.
p	an auxiliary probability for the crosswise question.
p.prime	an auxiliary probability for the anchor question.
data	a data frame containing information from the crosswise model and covariates.
start	Optional numeric vector of starting values for the beta and theta parameters. By default, starts are derived from binomial GLMs for the observed crosswise and anchor responses.
n.start	Number of optimization starts, including the data-informed start.
control	A list of control settings passed to [stats::optim()]. ‘fnscale‘ is fixed internally because the log-likelihood is maximized.

Value

An object of class ‘cmreg‘, a list with:

Call The matched model call.

Coefficients A coefficient table for the latent-trait outcome model.

AuxiliaryCoef A coefficient table for the anchor-response model.

VCV The estimated variance-covariance matrix of all parameters.

estimates, std.errors Named full-precision parameter estimates and standard errors.

logLik, n, convergence The optimized log-likelihood, number of complete observations, and optimizer convergence code.

References

Atsusaka, Y. and Stevenson, R. T. (2023). The crosswise model for sensitive survey questions. doi:10.1017/pan.2021.43.

See Also

[cmpredict()] for predicted latent-trait probabilities and [cmreg_p()] for a model with the latent trait as a predictor.

Examples

```
m <- cmreg(Y ~ female + age, anchor = A, p = 0.1, p.prime = 0.15,
           data = cmdata2)
m
```

`cmreg-methods` *Print and summarise crosswise regression fits*

Description

Print and summarise crosswise regression fits

Usage

```
## S3 method for class 'cmreg'
print(x, digits = max(3L, getOption("digits") - 3L), ...)

## S3 method for class 'cmreg'
summary(object, ...)

## S3 method for class 'summary.cmreg'
print(x, digits = max(3L, getOption("digits") - 3L), ...)
```

Arguments

`x`	A ‘cmreg‘ or ‘cmreg_p‘ fit object.
`digits`	Number of significant digits to display.
`...`	Additional arguments, currently ignored.
`object`	A ‘cmreg‘ or ‘cmreg_p‘ fit object.

Value

‘summary.cmreg()‘ returns a list containing the fit call, coefficient tables, log-likelihood, sample size, convergence code, and model type.

cmreg_p *cmreg_p*

Description

cmreg is used to run a regression with the latent sensitive trait as a predictor.

Usage

```
cmreg_p(
  formula,
  crosswise,
  anchor,
  p,
  p.prime,
  data,
  start = NULL,
  n.start = 3L,
  control = list()
)

cmreg.p(...)
```

Arguments

formula	an object of class "formula": a symbolic description of the model to be fitted. Ex. Outcome ~ Covariates. The crosswise and anchor responses are supplied separately through ‘crosswise‘ and ‘anchor‘.
crosswise	Unquoted column name (or a single character column name) for the crosswise response.
anchor	Unquoted column name (or a single character column name) for the anchor response.
p	an auxiliary probability for the crosswise question.
p.prime	an auxiliary probability for the anchor question.
data	a data frame containing information from the crosswise model, the outcome variable, and covariates.
start	Optional numeric vector of starting values on the reporting scale: beta, theta, gamma, crosswise effect, and a positive sigma. By default, starts are derived from GLM and least-squares fits to observed responses.
n.start	Number of optimization starts, including the data-informed start.
control	A list of control settings passed to [stats::optim()]. ‘fnscale‘ is fixed internally because the log-likelihood is maximized.
...	Arguments passed to cmreg_p().

Value

An object of class ‘cmreg_p‘ (and ‘cmreg‘), a list with:

Call The matched model call.

Coefficients A coefficient table for the Gaussian outcome model, including the latent-trait effect.

AuxiliaryCoef, AuxiliaryCoef2 Coefficient tables for the crosswise and anchor-response models, with the residual standard deviation.

VCV The estimated variance-covariance matrix of all parameters.

estimates, std.errors Named full-precision parameter estimates and standard errors.

logLik, n, convergence The optimized log-likelihood, number of complete observations, and optimizer convergence code.

References

Atsusaka, Y. and Stevenson, R. T. (2023). The crosswise model for sensitive survey questions. doi:10.1017/pan.2021.43.

See Also

[cmpredict_p()] for predictions conditional on latent-trait status and [cmreg()] for a model with the latent trait as the outcome.

Examples

```
# Keep the fitted example small: this likelihood is deliberately
# computationally intensive and package examples run during CRAN checks.
example_data <- cmdata3[seq_len(30L), ]
m2 <- cmreg_p(V ~ age + female, crosswise = Y, anchor = A, p = 0.1,
              p.prime = 0.15, data = example_data, n.start = 1L)
m2
```

`sim_cwdata` *Simulate Survey Data and Compute Bias-Corrected Estimates*

Description

Core simulation function that generates crosswise model data with inattentive respondents and computes bias-corrected prevalence estimates with bootstrap confidence intervals. This implements the methodology described in Appendix C5 for power analysis of the crosswise model.

Usage

```
sim_cwdata(
  N.sim = 500,
  sample,
  prevalence,
  p,
  p.prime,
  gamma,
  direct,
  verbose = TRUE
)

sim.cwdata(...)
```

Arguments

N.sim	Integer. Number of Monte Carlo simulations to run. Default is 500.
sample	Integer. Sample size per simulation (number of respondents).
prevalence	Numeric. True prevalence rate of the sensitive attribute (between 0 and 1).
p	Numeric. Probability for the randomization item in sensitive question (between 0 and 1).
p.prime	Numeric. Probability for the anchor question (non-sensitive, between 0 and 1).
gamma	Numeric. Proportion of attentive respondents (between 0 and 1).
direct	Numeric. Direct questioning estimate for comparison purposes (between 0 and 1).
verbose	Logical. If 'TRUE' (the default), display a progress bar while simulations run.
...	Arguments passed to sim_cwdata().

Details

The function implements the crosswise model with bias correction for inattentive respondents. For each simulation:

- Generates attentive/inattentive status based on gamma
- Simulates responses to sensitive question (crosswise format)
- Simulates responses to anchor question (for estimating attention rate)
- Computes naive crosswise estimate
- Applies bias correction based on estimated inattention
- Uses bootstrap (500 iterations) to compute 95% confidence intervals

The bias correction formula is:

$$\hat{\pi}_{BC} = \hat{\pi}_{naive} - \hat{Bias}$$

where the bias is estimated using the anchor question responses.

Value

A list containing:

Results Named vector with summary statistics including average bias, RMSE, and coverage rates for both naive and bias-corrected estimators

BiasCorrectEst Numeric vector of bias-corrected point estimates (sorted)

BiasCorrectLow Numeric vector of lower bounds of 95% bootstrap CIs (sorted)

BiasCorrectHigh Numeric vector of upper bounds of 95% bootstrap CIs (sorted)

EstimatedBias Numeric vector of estimated bias values for each simulation

RelativeLengthCI Numeric vector of relative CI lengths (bias-corrected vs naive)

References

Atsusaka and Stevenson (2023). Appendix C5: Sample Size Determination and Parameter Selection. doi:10.1017/pan.2021.43.

Examples

```
# Basic usage
result <- sim_cwdata(
  N.sim = 100,
  sample = 500,
  prevalence = 0.1,
  p = 0.1,
  p.prime = 0.1,
  gamma = 0.8,
  direct = 0.05
)
print(result$Results)
```

`sim_estimates` *Simulate and Plot Panel C of Figure C7*

Description

Runs Monte Carlo simulations using the bias-corrected crosswise model and creates a caterpillar plot showing sorted point estimates with bootstrap confidence intervals, replicating Panel C of Figure C7 in Atsusaka and Stevenson (2023).

Usage

```
sim_estimates(
  N.sim = 100,
  sample,
  prevalence,
  p,
  p.prime,
  gamma,
  direct,
  txcol = "dimgray",
  sim.results = NULL,
  verbose = TRUE
)

sim.estimates(...)
```

Arguments

N.sim	Integer. Number of Monte Carlo simulations. Default is 100.
sample	Integer. Sample size per simulation.
prevalence	Numeric. True prevalence rate of the sensitive attribute.
p	Numeric. Randomization probability for the sensitive question.
p.prime	Numeric. Randomization probability for the anchor question.
gamma	Numeric. Proportion of attentive respondents (between 0 and 1).
direct	Numeric. Direct questioning estimate for comparison.
txcol	Character. Color for annotation text. Default: "dimgray".
sim.results	Optional list. Pre-computed output from sim_cwdata. If NULL (default), the simulation is run internally.
verbose	Logical. Passed to sim_cwdata() when a new simulation is needed. If TRUE (the default), a progress bar is displayed.
...	Arguments passed to sim_estimates().

Details

The plot displays:

- Sorted bias-corrected point estimates as filled circles
- Bootstrap 95% confidence intervals as vertical line segments
- A horizontal reference line at 0
- A horizontal reference line at the true prevalence prevalence (red)

Estimates are sorted in ascending order, creating a characteristic "fan" shape that reveals the distribution of estimates across simulations.

Text annotation positions are calibrated to N.sim = 100 and scale proportionally for other values.

If sim.results is supplied, all simulation parameters (N.sim, sample, p, p.prime, gamma, direct) are still used for the annotations, but no new simulation is run.

Value

Invisibly returns the simulation results list from `sim_cwdata`, containing `BiasCorrectEst`, `BiasCorrectLow`, `BiasCorrectHigh`, and summary `Results`.

References

Atsusaka and Stevenson (2023). Figure C7, Panel C. doi:10.1017/pan.2021.43.

See Also

`sim_cwdata` for the underlying simulation function

Examples

```
# Replicate Panel C of Figure C7
sim_estimates(
  N.sim   = 100,
  sample  = 1000,
  prevalence = 0.1,
  p       = 0.1,
  p.prime = 0.1,
  gamma   = 0.8,
  direct  = 0.1
)

# Re-use pre-computed simulation results
res <- sim_cwdata(N.sim = 100, sample = 1000, prevalence = 0.1,
                  p = 0.1, p.prime = 0.1, gamma = 0.8, direct = 0.1)
sim_estimates(sample = 1000, prevalence = 0.1, p = 0.1, p.prime = 0.1,
              gamma = 0.8, direct = 0.1, sim.results = res)
```

---

sim_power *Compute Statistical Power for a Fixed Sample Size*

Description

Computes statistical power for a one-sided hypothesis test of H0: pi <= pi0 versus H1: pi > pi0 at a given fixed sample size. This is useful when researchers already know their sample size and want to assess the power they can expect from their study.

Usage

```
sim_power(
  N.sim,
  sample,
  pi.null,
```

```
  pi.alt,
  p,
  p.prime,
  gamma,
  direct,
  verbose = TRUE
)

sim.power(...)
```

Arguments

N.sim	Integer. Number of Monte Carlo simulations. Larger values provide more stable power estimates but increase computation time.
sample	Integer. The fixed sample size (number of respondents) to evaluate.
pi.null	Numeric. Prevalence rate under the null hypothesis (pi0). Can be 0 or a value from direct questioning.
pi.alt	Numeric. True prevalence rate under the alternative hypothesis (pi1). Must be greater than pi.null.
p	Numeric. Probability for the randomization item in the sensitive question. Values between 0.1 and 0.3 are typical.
p.prime	Numeric. Probability for the anchor question (non-sensitive).
gamma	Numeric. Proportion of attentive respondents (between 0 and 1). For example, 0.8 means 80% of respondents are attentive.
direct	Numeric. Direct questioning estimate for comparison purposes.
verbose	Logical. If TRUE (the default), report progress messages and show progress bars for the two underlying simulations.
...	Arguments passed to sim_power().

Details

The function implements the power calculation based on the Wald test:

$$Power = \Phi\left(\frac{\pi_1 - \pi_0 + z_\alpha \tilde{\sigma}_0}{\tilde{\sigma}_1}\right)$$

where:

- Φ is the cumulative distribution function of the standard normal
- z_α is derived from the simulated null distribution
- $\tilde{\sigma}_0$ and $\tilde{\sigma}_1$ are simulated standard errors under H0 and H1 respectively

The function runs sim_cwdata() twice: once under H0 using pi.null and once under H1 using pi.alt, to estimate the sampling distributions at the specified sample size.

Value

A numeric scalar representing the estimated statistical power (probability of correctly rejecting H0 when H1 is true) at the given sample size.

Note

This function can take considerable time to run depending on N.sim and sample size. For quick exploration, use smaller N.sim values (for example, 100 to 500). For publication-quality results, use N.sim >= 2000.

References

Atsusaka and Stevenson (2023). Appendix C5: Sample Size Determination and Parameter Selection. doi:10.1017/pan.2021.43.

See Also

sim_cwdata for the underlying simulation function

Examples

```
# Compute power at a fixed sample size of 1000

pwr <- sim_power(
  N.sim  = 500,
  sample = 1000,
  pi.null = 0,
  pi.alt  = 0.1,
  p       = 0.1,
  p.prime = 0.1,
  gamma   = 0.8,
  direct  = 0.02
)
cat(sprintf("Estimated power: %.3f\n", pwr))
```

---

sim_power_N *Determine Sample Size for Desired Coverage Properties*

Description

Tests multiple sample sizes to determine which achieves desired confidence interval coverage properties. Specifically, it calculates what percentage of 95% confidence intervals (1) exclude zero and (2) include the direct estimate. This helps researchers choose a sample size that provides sufficient precision.

Usage

```
sim_power_N(N.sim = 50, prevalence, p, p.prime, gamma, direct, verbose = TRUE)

sim.power.N(...)
```

Arguments

N.sim	Integer. Number of Monte Carlo simulations per sample size. Default is 50. Larger values provide more stable estimates but increase computation time.
prevalence	Numeric. True prevalence rate of the sensitive attribute (between 0 and 1).
p	Numeric. Probability for the randomization item in sensitive question.
p.prime	Numeric. Probability for the anchor question (non-sensitive).
gamma	Numeric. Proportion of attentive respondents (between 0 and 1).
direct	Numeric. Direct questioning estimate for comparison purposes.
verbose	Logical. If TRUE (the default), display a progress bar while sample sizes are evaluated.
...	Arguments passed to sim_power_N().

Details

This function is useful for planning studies where researchers want to:

1. Distinguish the estimated prevalence from zero with high confidence
2. Obtain narrow confidence intervals for precise estimation
3. Compare crosswise estimates with direct questioning estimates

For each sample size, the function:

- Simulates N.sim datasets using the crosswise model
- Computes bias-corrected estimates with bootstrap 95% CIs
- Calculates what percentage of CIs contain zero
- Calculates what percentage of CIs contain the direct estimate

A progress bar displays simulation progress.

Value

A data frame with three columns:

SampleSize Sample sizes tested: 100, 500, 1000, 1500, 2000, 2500, 3000

CoverageZero Percentage of 95% CIs that include zero. Lower values indicate better precision (CIs exclude zero).

CoverageDirect Percentage of 95% CIs that include the direct estimate. Values near 95% suggest good agreement with direct questioning.

Note

- Low CoverageZero values indicate CIs that reliably exclude zero (good for establishing that prevalence is non-zero)
- CoverageDirect near 95% suggests consistency between crosswise and direct questioning approaches
- This function uses 500 bootstrap iterations per simulation

References

Atsusaka and Stevenson (2023). Appendix C5: Sample Size Determination and Parameter Selection. doi:10.1017/pan.2021.43.

See Also

sim_power for the underlying simulation function

Examples

```
# Find sample size needed to reliably exclude zero

result <- sim_power_N(
  N.sim = 50,
  prevalence = 0.1,
  p = 0.1,
  p.prime = 0.1,
  gamma = 0.8,
  direct = 0.05
)

print(result)

# Visualize results
plot(result$SampleSize, result$CoverageZero,
     type = "b", xlab = "Sample Size",
     ylab = "% of CIs Including Zero",
     main = "Precision vs Sample Size")
```

Index