Package ‘RRTCS’

November 3, 2015

Type Package

Title Randomized Response Techniques for Complex Surveys

Version 0.0.3

Date 2015-04-22

Author Beatriz Cobo Rodríguez, María del Mar Rueda García, Antonio Arcos Cebrián

Maintainer Beatriz Cobo Rodríguez <beacr@ugr.es>

Description Point and interval estimation of linear parameters with data obtained from complex surveys (including stratified and clustered samples) when randomization techniques are used. The randomized response technique was developed to obtain estimates that are more valid when studying sensitive topics. Estimators and variances for 14 randomized response methods for qualitative variables and 7 randomized response methods for quantitative variables are also implemented. In addition, some data sets from surveys with these randomization methods are included in the package.

License GPL (>= 2)

Encoding UTF-8

Imports sampling, samplingVarEst, stats

Suggests knitr

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2015-11-03 13:36:04

R topics documented:

RRTCS-package *Randomized Response Techniques for Complex Surveys*

Description

The aim of this package is to calculate point and interval estimation for linear parameters with data obtained from randomized response surveys. Twenty one RR methods are implemented for complex surveys:

- Randomized response procedures to estimate parameters of a qualitative stigmatizing characteristic: Christofides model, Devore model, Forced-Response model, Horvitz model, Horvitz model with unknown B, Kuk model, Mangat model, Mangat model with unknown B, Mangat-Singh model, Mangat-Singh-Singh model, Mangat-Singh-Singh model with unknown B, Singh-Joarder model, SoberanisCruz model and Warner model.

- Randomized response procedures to estimate parameters of a quantitative stigmatizing characteristic: BarLev model, Chaudhuri-Christofides model, Diana-Perri-1 model, Diana-Perri-2 model, Eichhorn-Hayre model, Eriksson model and Saha model.

Using the usual notation in survey sampling, we consider a finite population $U = \{1, \ldots, i, \ldots, N\}$, consisting of N different elements. Let y_i be the value of the sensitive aspect under study for the ith population element. Our aim is to estimate the finite population total $Y = \sum_{i=1}^N y_i$ of the variable of interest y or the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^N y_i$. If we can estimate the proportion of the population presenting a certain stigmatized behaviour A, the variable y_i takes the value 1 if $i \in G_A$ (the group with the stigmatized behaviour) and the value zero otherwise. Some qualitative models use an innocuous or related attribute B whose population proportion can be known or unknown.

Assume that a sample s is chosen according to a general design p with inclusion probabilities $\pi_i = \sum_{s \ni i} p(s), i \in U$.

In order to include a wide variety of RR procedures, we consider the unified approach given by Arnab (1994). The interviews of individuals in the sample s are conducted in accordance with the RR model. For each $i \in s$ the RR induces a random response z_i (denoted scrambled response) so that the revised randomized response r_i (Chaudhuri and Christofides, 2013) is an unbiased estimation of y_i. Then, an unbiased estimator for the population total of the sensitive characteristic y is given by

$$\widehat{Y}_R = \sum_{i \in s} \frac{r_i}{\pi_i}$$

The variance of this estimator is given by:

$$V(\widehat{Y}_R) = \sum_{i \in U} \frac{V_R(r_i)}{\pi_i} + V_{HT}(r)$$

where $V_R(r_i)$ is the variance of r_i under the randomized device and $V_{HT}(r)$ is the design-variance of the Horvitz Thompson estimator of r_i values.

This variance is estimated by:

$$\widehat{V}(\widehat{Y}_R) = \sum_{i \in s} \frac{\widehat{V}_R(r_i)}{\pi_i} + \widehat{V}(r)$$

where $\widehat{V}_R(r_i)$ varies with the RR device and the estimation of the design-variance, $\widehat{V}(r)$, is obtained using Deville's method (Deville, 1993).

The confidence interval at $(1-\alpha)$ % level is given by

$$ci = \left(\widehat{Y}_R - z_{1-\frac{\alpha}{2}}\sqrt{\widehat{V}(\widehat{Y}_R)}, \widehat{Y}_R + z_{1-\frac{\alpha}{2}}\sqrt{\widehat{V}(\widehat{Y}_R)}\right)$$

where $z_{1-\frac{\alpha}{2}}$ denotes the $(1-\alpha)$ % quantile of a standard normal distribution.

Similarly, an unbiased estimator for the population mean $\bar{Y}$ is given by

$$\widehat{\bar{Y}}_R = \frac{1}{N}\sum_{i\in s}\frac{r_i}{\pi_i}$$

and an unbiased estimator for its variance is calculated as:

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{1}{N^2}\left(\sum_{i\in s}\frac{\widehat{V}_R(r_i)}{\pi_i} + \widehat{V}(r)\right)$$

In cases where the population size N is unknown, we consider Hàjek-type estimators for the mean:

$$\widehat{\bar{Y}}_{RH} = \frac{\sum_{i\in s} r_i}{\sum_{i\in s}\frac{1}{\pi_i}}$$

and Taylor-series linearization variance estimation of the ratio (Wolter, 2007) is used.

In qualitative models, the values r_i and $\widehat{V}_R(r_i)$ for $i \in s$ are described in each model.

In some quantitative models, the values r_i and $\widehat{V}_R(r_i)$ for $i \in s$ are calculated in a general form (Arcos et al, 2015) as follows:

The randomized response given by the person i is

$$z_i = \begin{cases} y_i & \text{with probability } p_1 \\ y_i S_1 + S_2 & \text{with probability } p_2 \\ S_3 & \text{with probability } p_3 \end{cases}$$

with $p_1 + p_2 + p_3 = 1$ and where S_1, S_2 and S_3 are scramble variables whose distributions are assumed to be known. We denote by μ_i and σ_i respectively the mean and standard deviation of the variable $S_i, (i = 1, 2, 3)$.

The transformed variable is

$$r_i = \frac{z_i - p_2\mu_2 - p_3\mu_3}{p_1 + p_2\mu_1},$$

its variance is

$$V_R(r_i) = \frac{1}{(p_1 + p_2\mu_1)^2}(y_i^2 A + y_i B + C)$$

where

$$A = p_1(1-p_1) + \sigma_1^2 p_2 + \mu_1^2 p_2 - \mu_1^2 p_2^2 - 2p_1p_2\mu_1$$

$$B = 2p_2\mu_1\mu_2 - 2\mu_1\mu_2p_2^2 - 2p_1p_2\mu_2 - 2\mu_3p_1p_3 - 2\mu_1\mu_3p_2p_3$$

$$C = (\sigma_2^2 + \mu_2^2)p_2 + (\sigma_3^2 + \mu_3^2)p_3 - (\mu_2p_2 + \mu_3p_3)^2$$

and the estimated variance is

$$\widehat{V}_R(r_i) = \frac{1}{(p_1 + p_2\mu_1)^2}(r_i^2 A + r_i B + C).$$

Some of the quantitative techniques considered can be viewed as particular cases of the above described procedure. Other models are described in the respective function.

Alternatively, the variance can be estimated using certain resampling methods.

Author(s)

Beatriz Cobo Rodríguez, Department of Statistics and Operations Research. University of Granada <beacr@ugr.es>

María del Mar Rueda García, Department of Statistics and Operations Research. University of Granada <mrueda@ugr.es>

Antonio Arcos Cebrián, Department of Statistics and Operations Research. University of Granada <arcos@ugr.es>

Maintainer: Beatriz Cobo Rodríguez <beacr@ugr.es>

References

Arcos, A., Rueda, M., Singh, S. (2015). *A generalized approach to randomised response for quantitative variables.* Quality and Quantity 49, 1239-1256.

Arnab, R. (1994). *Non-negative variance estimator in randomized response surveys.* Comm. Stat. Theo. Math. 23, 1743-1752.

Chaudhuri, A., Christofides, T.C. (2013). *Indirect Questioning in Sample Surveys* Springer-Verlag Berlin Heidelberg.

Deville, J.C. (1993). *Estimation de la variance pour les enquêtes en deux phases.* Manuscript, INSEE, Paris.

Wolter, K.M. (2007). *Introduction to Variance Estimation.* 2nd Edition. Springer.

`BarLev` *BarLev model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the BarLev model. The function can also return the transformed variable. The BarLev model was proposed by Bar-Lev et al. in 2004.

Usage

```
BarLev(z,p,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	probability of direct response
`mu`	mean of the scramble variable S
`sigma`	standard deviation of the scramble variable S
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person i is

$$z_i = \begin{cases} y_i & \text{with probability } p \\ y_i S & \text{with probability } 1-p \end{cases}$$

where S is a scramble variable, whose mean μ and standard deviation σ are known.

Value

Point and confidence estimates of the sensitive characteristics using the BarLev model. The transformed variable is also reported, if required.

References

Bar-Lev S.K., Bobovitch, E., Boukai, B. (2004). *A note on randomized response models for quantitative data.* Metrika, 60, 255-260.

See Also

BarLevData

ResamplingVariance

Examples

```
data(BarLevData)
dat=with(BarLevData,data.frame(z,Pi))
p=0.6
mu=1
sigma=1
cl=0.95
BarLev(dat$z,p,mu,sigma,dat$Pi,"total",cl)
```

`BarLevData` *Randomized Response Survey on industrial company income*

Description

This data set contains observations from a randomized response survey conducted in a population of 2396 industrial companies in a city to investigate their income. The sample is drawn by stratified sampling with probabilities proportional to the size of the company. The randomized response technique used is the BarLev model (Bar-Lev et al, 2004) with parameter $p = 0.6$ and scramble variable $S = exp(1)$.

Usage

```
BarLevData
```

Format

A data frame containing 370 observations of a sample of companies divided into three strata. The variables are:

- ID: Survey ID
- ST: Strata ID
- z: The randomized response to the question: What was the company's income in the previous fiscal year?
- Pi: first-order inclusion probabilities

References

Bar-Lev S.K., Bobovitch, E., Boukai, B. (2004). *A note on randomized response models for quantitative data.* Metrika, 60, 255-260.

See Also

`BarLev`

Examples

```
data(BarLevData)
```

`ChaudhuriChristofides` *Chaudhuri-Christofides model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Chaudhuri-Christofides model. The function can also return the transformed variable. The Chaudhuri-Christofides model can be seen in Chaudhuri and Christofides (2013, page 97).

Usage

```
ChaudhuriChristofides(z,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`mu`	vector with the means of the scramble variables
`sigma`	vector with the standard deviations of the scramble variables
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person i is $z_i = y_i S_1 + S_2$ where S_1, S_2 are scramble variables, whose mean μ and standard deviation σ are known.

Value

Point and confidence estimates of the sensitive characteristics using the Chaudhuri-Christofides model. The transformed variable is also reported, if required.

References

Chaudhuri, A., and Christofides, T.C. (2013) *Indirect Questioning in Sample Surveys.* Springer-Verlag Berlin Heidelberg.

See Also

`ChaudhuriChristofidesData`

`ChaudhuriChristofidesDatapij`

`ResamplingVariance`

Examples

```
N=417
data(ChaudhuriChristofidesData)
dat=with(ChaudhuriChristofidesData,data.frame(z,Pi))
mu=c(6,6)
sigma=sqrt(c(10,10))
cl=0.95
data(ChaudhuriChristofidesDatapij)
ChaudhuriChristofides(dat$z,mu,sigma,dat$Pi,"mean",cl,pij=ChaudhuriChristofidesDatapij)
```

ChaudhuriChristofidesData *Randomized Response Survey on agricultural subsidies*

Description

This data set contains observations from a randomized response survey conducted in a population of 417 individuals in a municipality to investigate the agricultural subsidies. The sample is drawn by sampling with unequal probabilities (probability proportional to agricultural subsidies in the previous year). The randomized response technique used is the Chaudhuri-Christofides model (Chaudhuri and Christofides, 2013) with scramble variables $S_1 = U(1, ..., 11)$ and $S_2 = U(1, ..., 11)$.

Usage

```
ChaudhuriChristofidesData
```

Format

A data frame containing 100 observations. The variables are:

- ID: Survey ID
- z: The randomized response to the question: What are your annual agricultural subsidies?
- Pi: first-order inclusion probabilities

References

Chaudhuri, A., and Christofides, T.C. (2013) *Indirect Questioning in Sample Surveys.* Springer-Verlag Berlin Heidelberg.

See Also

ChaudhuriChristofides

ChaudhuriChristofidesDatapij

Examples

```
data(ChaudhuriChristofidesData)
```

ChaudhuriChristofidesDatapij *Matrix of the second-order inclusion probabilities*

Description

This dataset consists of a square matrix of dimension 100 with the first and second order inclusion probabilities for the units included in sample s, drawn from a population of size $N = 417$ according to a sampling with unequal probabilities (probability proportional to agricultural subsidies in the previous year).

Usage

```
ChaudhuriChristofidesDatapij
```

See Also

ChaudhuriChristofides

ChaudhuriChristofidesData

Examples

```
data(ChaudhuriChristofidesDatapij)
#Now, let select only the first-order inclusion probabilities
diag(ChaudhuriChristofidesDatapij)
```

Christofides *Christofides model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Christofides model. The function can also return the transformed variable. The Christofides model was proposed by Christofides in 2003.

Usage

```
Christofides(z,mm,pm,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`mm`	vector with the marks of the cards
`pm`	vector with the probabilities of previous marks
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Christofides randomized response technique, a sampled person i is given a box with identical cards, each bearing a separate mark as $1, \ldots, k, \ldots m$ with $m \geq 2$ but in known proportions $p_1, \ldots, p_k, \ldots p_m$ with $0 < p_k < 1$ for $k = 1, \ldots, m$ and $\sum_{k=1}^{m} p_k = 1$. The person sampled is requested to draw one of the cards and respond

$$z_i = \begin{cases} k & \text{if a card marked } k \text{ is drawn and the person bears } A^c \\ m-k+1 & \text{if a card marked } k \text{ is drawn but the person bears } A \end{cases}$$

The transformed variable is $r_i = \frac{z_i - \mu}{m+1-2\mu}$ where $\mu = \sum_{k=1}^{m} k p_k$ and the estimated variance is $\widehat{V}_R(r_i) = \frac{V_R(k)}{(m+1-2\mu)^2}$, where $V_R(k) = \sum_{k=1}^{m} k^2 p_k - \mu^2$.

Value

Point and confidence estimates of the sensitive characteristics using the Christofides model. The transformed variable is also reported, if required.

References

Christofides, T.C. (2003). *A generalized randomized response technique.* Metrika, 57, 195-200.

See Also

`ChristofidesData`

`ResamplingVariance`

Examples

```
N=802
data(ChristofidesData)
dat=with(ChristofidesData,data.frame(z,Pi))
mm=c(1,2,3,4,5)
pm=c(0.1,0.2,0.3,0.2,0.2)
cl=0.95
Christofides(dat$z,mm,pm,dat$Pi,"mean",cl,N)
```

`ChristofidesData` *Randomized Response Survey on eating disorders*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate eating disorders. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Christofides model (Christofides, 2003) with parameters, $mm = (1, 2, 3, 4, 5)$ and $pm = (0.1, 0.2, 0.3, 0.2, 0.2)$.

Usage

```
ChristofidesData
```

Format

A data frame containing 150 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you have problems of anorexia or bulimia?
- Pi: first-order inclusion probabilities

References

Christofides, T.C. (2003). *A generalized randomized response technique.* Metrika, 57, 195-200.

See Also

`Christofides`

Examples

```
data(ChristofidesData)
```

`Devore` *Devore model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Devore model. The function can also return the transformed variable. The Devore model was proposed by Devore in 1977.

Usage

```
Devore(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of cards bearing the mark A
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Devore model, the randomized response device presents to the sampled person labelled i a box containing a large number of identical cards with a proportion p, $(0 < p < 1)$ bearing the mark A and the rest marked B (an innocuous attribute). The response solicited denoted by z_i takes the value y_i if i bears A and the card drawn is marked A. Otherwise z_i takes the value 1.

The transformed variable is $r_i = \frac{z_i-(1-p)}{p}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Devore model. The transformed variable is also reported, if required.

References

Devore, J.L. (1977). *A note on the randomized response technique.* Communications in Statistics Theory and Methods 6: 1525-1529.

See Also

DevoreData

ResamplingVariance

Examples

```
data(DevoreData)
dat=with(DevoreData,data.frame(z,Pi))
p=0.7
cl=0.95
Devore(dat$z,p,dat$Pi,"total",cl)
```

DevoreData *Randomized Response Survey on instant messaging*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate the use of instant messaging. The sample is drawn by stratified sampling by academic year. The randomized response technique used is the Devore model (Devore, 1977) with parameter $p = 0.7$. The unrelated question is: Are you alive?

Usage

```
DevoreData
```

Format

A data frame containing 240 observations divided into four strata. The sample is selected from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- z: The randomized response to the question: Do you use whatsapp / line or similar instant messaging while you study?
- Pi: first-order inclusion probabilities

References

Devore, J.L. (1977). *A note on the randomized response technique.* Communications in Statistics Theory and Methods 6: 1525-1529.

See Also

Devore

Examples

```
data(DevoreData)
```

`DianaPerri1` *Diana-Perri-1 model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Diana-Perri-1 model. The function can also return the transformed variable. The Diana-Perri-1 model was proposed by Diana and Perri (2010, page 1877).

Usage

```
DianaPerri1(z,p,mu,pi,type=c("total","mean"),cl,N=NULL,method="srswr")
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	probability of direct response
`mu`	vector with the means of the scramble variables W and U
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`method`	method used to draw the sample: srswr or srswor. By default it is srswr

Details

In the Diana-Perri-1 model let $p \in [0,1]$ be a design parameter, controlled by the experimenter, which is used to randomize the response as follows: with probability p the interviewer responds with the true value of the sensitive variable, whereas with probability $1-p$ the respondent gives a coded value, $z_i = W(y_i + U)$ where W, U are scramble variables whose distribution is assumed to be known.

To estimate $\bar{Y}$ a sample of respondents is selected according to simple random sampling with replacement. The transformed variable is

$$r_i = \frac{z_i - (1-p)\mu_W \mu_U}{p + (1-p)\mu_W}$$

where μ_W, μ_U are the means of W, U scramble variables, respectively.

The estimated variance in this model is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n(p + (1-p)\mu_W)^2}$$

where $s_z^2 = \sum_{i=1}^n \frac{(z_i - \bar{z})^2}{n-1}$.

If the sample is selected by simple random sampling without replacement, the estimated variance is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n(p + (1-p)\mu_W)^2} \left(1 - \frac{n}{N}\right)$$

Value

Point and confidence estimates of the sensitive characteristics using the Diana-Perri-1 model. The transformed variable is also reported, if required.

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875-1890.

See Also

DianaPerri1Data

DianaPerri2

ResamplingVariance

Examples

```
N=417
data(DianaPerri1Data)
dat=with(DianaPerri1Data,data.frame(z,Pi))
p=0.6
mu=c(5/3,5/3)
cl=0.95
DianaPerri1(dat$z,p,mu,dat$Pi,"mean",cl,N,"srswor")
```

DianaPerri1Data *Randomized Response Survey on defrauded taxes*

Description

This data set contains observations from a randomized response survey conducted in a population of 417 individuals in a municipality to investigate defrauded taxes. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Diana and Perri 1 model (Diana and Perri, 2010) with parameters $p = 0.6$, $W = F(10, 5)$ and $U = F(5, 5)$.

Usage

```
DianaPerri1Data
```

Format

A data frame containing 150 observations from a population of $N = 417$. The variables are:

- ID: Survey ID
- z: The randomized response to the question: What quantity of your agricultural subsidy do you declare in your income tax return?
- Pi: first-order inclusion probabilities

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875–1890.

See Also

DianaPerri1

Examples

```
data(DianaPerri1Data)
```

`DianaPerri2` *Diana-Perri-2 model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Diana-Perri-2 model. The function can also return the transformed variable. The Diana-Perri-2 model was proposed by Diana and Perri (2010, page 1879).

Usage

```
DianaPerri2(z,mu,beta,pi,type=c("total","mean"),cl,N=NULL,method="srswr")
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`mu`	vector with the means of the scramble variables W and U
`beta`	the constant of weighting
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`method`	method used to draw the sample: srswr or srswor. By default it is srswr

Details

In the Diana-Perri-2 model, each respondent is asked to report the scrambled response $z_i = W(\beta U + (1-\beta)y_i)$ where $\beta \in [0,1)$ is a suitable constant controlled by the researcher and W, U are scramble variables whose distribution is assumed to be known.

To estimate $\bar{Y}$ a sample of respondents is selected according to simple random sampling with replacement. The transformed variable is

$$r_i = \frac{z_i - \beta\mu_W\mu_U}{(1-\beta)\mu_W}$$

where μ_W, μ_U are the means of W, U scramble variables, respectively.

The estimated variance in this model is

$$\widehat{V}(\widehat{\widehat{Y}}_R) = \frac{s_z^2}{n(1-\beta)^2\mu_W^2}$$

where $s_z^2 = \sum_{i=1}^{n} \frac{(z_i-\bar{z})^2}{n-1}$.

If the sample is selected by simple random sampling without replacement, the estimated variance is

$$\widehat{V}(\widehat{\widehat{Y}}_R) = \frac{s_z^2}{n(1-\beta)^2\mu_W^2}\left(1-\frac{n}{N}\right)$$

Value

Point and confidence estimates of the sensitive characteristics using the Diana-Perri-2 model. The transformed variable is also reported, if required.

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875-1890.

See Also

DianaPerri2Data

DianaPerri1

ResamplingVariance

Examples

```
N=100000
data(DianaPerri2Data)
dat=with(DianaPerri2Data,data.frame(z,Pi))
beta=0.8
mu=c(50/48,5/3)
cl=0.95
DianaPerri2(dat$z,mu,beta,dat$Pi,"mean",cl,N,"srswor")
```

DianaPerri2Data *Randomized Response Survey of a simulated population*

Description

This data set contains observations from a simulated randomized response survey. The interest variable is a normal distribution with mean 1500 and standard deviation 4. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Diana and Perri 2 model (Diana and Perri, 2010) with parameters $W = F(10,50), U = F(1,5)$ and $\beta = 0.8$.

Usage

```
DianaPerri2Data
```

Format

A data frame containing 1000 observations from a population of $N = 100000$. The variables are:

- ID: Survey ID
- z: The randomized response
- Pi: first-order inclusion probabilities

References

Diana, G., Perri, P.F. (2010). *New scrambled response models for estimating the mean of a sensitive quantitative character.* Journal of Applied Statistics 37 (11), 1875–1890.

See Also

DianaPerri2

Examples

```
data(DianaPerri2Data)
```

EichhornHayre *Eichhorn-Hayre model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Eichhorn-Hayre model. The function can also return the transformed variable. The Eichhorn-Hayre model was proposed by Eichhorn and Hayre in 1983.

Usage

```
EichhornHayre(z,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
mu	mean of the scramble variable S
sigma	standard deviation of the scramble variable S
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person labelled i is $z_i = y_i S$ where S is a scramble variable whose distribution is assumed to be known.

Value

Point and confidence estimates of the sensitive characteristics using the Eichhorn-Hayre model. The transformed variable is also reported, if required.

References

Eichhorn, B.H., Hayre, L.S. (1983). *Scrambled randomized response methods for obtaining sensitive quantitative data.* Journal of Statistical Planning and Inference, 7, 306-316.

See Also

EichhornHayreData

ResamplingVariance

Examples

```
data(EichhornHayreData)
dat=with(EichhornHayreData,data.frame(z,Pi))
mu=1.111111
sigma=0.5414886
cl=0.95
#This line returns a warning showing why the variance estimation is not possible.
#See ResamplingVariance for several alternatives.
EichhornHayre(dat$z,mu,sigma,dat$Pi,"mean",cl)
```

EichhornHayreData *Randomized Response Survey on family income*

Description

This data set contains observations from a randomized response survey conducted in a population of families to investigate their income. The sample is drawn by stratified sampling by house ownership. The randomized response technique used is the Eichhorn and Hayre model (Eichhorn and Hayre, 1983) with scramble variable $S = F(20, 20)$.

Usage

```
EichhornHayreData
```

Format

A data frame containing 150 observations of a sample extracted from a population of families divided into two strata. The variables are:

- ID: Survey ID
- ST: Strata ID
- z: The randomized response to the question: What is the annual household income?
- Pi: first-order inclusion probabilities

References

Eichhorn, B.H., Hayre, L.S. (1983). *Scrambled randomized response methods for obtaining sensitive quantitative data.* Journal of Statistical Planning and Inference, 7, 306-316.

See Also

EichhornHayre

Examples

```
data(EichhornHayreData)
```

Eriksson *Eriksson model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Eriksson model. The function can also return the transformed variable. The Eriksson model was proposed by Eriksson in 1973.

Usage

```
Eriksson(z,p,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	probability of direct response
mu	mean of the scramble variable S
sigma	standard deviation of the scramble variable S
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The randomized response given by the person labelled i is y_i with probability p and a discrete uniform variable S with probabilities $q_1, q_2, ..., q_j$ verifying $q_1 + q_2 + ... + q_j = 1 - p$.

Value

Point and confidence estimates of the sensitive characteristics using the Eriksson model. The transformed variable is also reported, if required.

References

Eriksson, S.A. (1973). *A new model for randomized response.* International Statistical Review 41, 40-43.

See Also

EriksssonData

ResamplingVariance

Examples

```
N=53376
data(EriksssonData)
dat=with(EriksssonData,data.frame(z,Pi))
p=0.5
mu=mean(c(0,1,3,5,8))
sigma=sqrt(4/5*var(c(0,1,3,5,8)))
cl=0.95
Eriksson(dat$z,p,mu,sigma,dat$Pi,"mean",cl,N)
```

EriksssonData *Randomized Response Survey on student cheating*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate cheating behaviour in exams. The sample is drawn by stratified sampling by university faculty with uniform allocation. The randomized response technique used is the Eriksson model (Eriksson, 1973) with parameter $p = 0.5$ and S a discrete uniform variable at the points (0,1,3,5,8).

The data were used by Arcos et al. (2015).

Usage

```
EriksssonData
```

Format

A data frame containing 102 students of a sample extracted from a population of $N = 53376$ divided into four strata. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- z: The randomized response to the question: How many times have you cheated in an exam in the past year?
- Pi: first-order inclusion probabilities

References

Arcos, A., Rueda, M. and Singh, S. (2015). *A generalized approach to randomised response for quantitative variables.* Quality and Quantity 49, 1239-1256.

Eriksson, S.A. (1973). *A new model for randomized response.* International Statistical Review 41, 40-43.

See Also

Eriksson

Examples

```
data(ErikssonData)
```

`ForcedResponse` *Forced-Response model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Forced-Response model. The function can also return the transformed variable. The Forced-Response model was proposed by Boruch in 1972.

Usage

```
ForcedResponse(z,p1,p2,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p1	proportion of cards marked "Yes"
p2	proportion of cards marked "No"
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean

`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Forced-Response scheme, the sampled person i is offered a box with cards: some are marked "Yes" with a proportion p_1, some are marked "No" with a proportion p_2 and the rest are marked "Genuine", in the remaining proportion $p_3 = 1-p_1-p_2$, where $0 < p_1, p_2 < 1, p_1 \neq p_2, p_1+p_2 < 1$. The person is requested to randomly draw one of them, to observe the mark on the card, and to respond

$$z_i = \begin{cases} 1 & \text{if the card is type "Yes"} \\ 0 & \text{if the card is type "No"} \\ y_i & \text{if the card is type "Genuine"} \end{cases}$$

The transformed variable is $r_i = \frac{z_i - p_1}{1-p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Forced-Response model. The transformed variable is also reported, if required.

References

Boruch, R.F. (1972). *Relations among statistical methods for assuring confidentiality of social research data.* Social Science Research, 1, 403-414.

See Also

ForcedResponseData

ForcedResponseDataSt

ResamplingVariance

Examples

```
data(ForcedResponseData)
dat=with(ForcedResponseData,data.frame(z,Pi))
p1=0.2
p2=0.2
cl=0.95
ForcedResponse(dat$z,p1,p2,dat$Pi,"total",cl)

#Forced Response with strata
data(ForcedResponseDataSt)
dat=with(ForcedResponseDataSt,data.frame(ST,z,Pi))
p1=0.2
p2=0.2
cl=0.95
ForcedResponse(dat$z,p1,p2,dat$Pi,"total",cl)
```

ForcedResponseData *Randomized Response Survey of a simulated population*

Description

This data set contains observations from a randomized response survey obtained from a simulated population. The main variable is a binomial distribution with a probability 0.5. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Forced Response model (Boruch, 1972) with parameters $p_1 = 0.2$ and $p_2 = 0.2$.

Usage

```
ForcedResponseData
```

Format

A data frame containing 1000 observations from a population of $N = 10000$. The variables are:

- ID: Survey ID
- z: The randomized response
- Pi: first-order inclusion probabilities

References

Boruch, R.F. (1972). *Relations among statistical methods for assuring confidentiality of social research data.* Social Science Research, 1, 403-414.

See Also

ForcedResponse

Examples

```
data(ForcedResponseData)
```

ForcedResponseDataSt *Randomized Response Survey on infertility*

Description

This data set contains observations from a randomized response survey to determine the prevalence of infertility among women of childbearing age in a population-base study. The sample is drawn by stratified sampling. The randomized response technique used is the Forced Response model (Boruch, 1972) with parameters $p_1 = 0.2$ and $p_2 = 0.2$.

Usage

```
ForcedResponseDataSt
```

Format

A data frame containing 442 observations. The variables are:

- ID: Survey ID
- ST: Strata ID
- z: The randomized response to the question: Did you ever have some medical treatment for the infertility?
- Pi: first-order inclusion probabilities

References

Boruch, R.F. (1972). *Relations among statistical methods for assuring confidentiality of social research data.* Social Science Research, 1, 403-414.

See Also

ForcedResponse

Examples

```
data(ForcedResponseDataSt)
```

Horvitz *Horvitz model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Horvitz model. The function can also return the transformed variable. The Horvitz model was proposed by Horvitz et al. (1967) and by Greenberg et al. (1969).

Usage

```
Horvitz(z,p,alpha,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive question
alpha	proportion of people with the innocuous attribute
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean

`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Horvitz model, the randomized response device presents to the sampled person labelled i a box containing a large number of identical cards, with a proportion p, $(0 < p < 1)$ bearing the mark A and the rest marked B (an innocuous attribute whose population proportion α is known). The response solicited denoted by z_i takes the value y_i if i bears A and the card drawn is marked A or if i bears B and the card drawn is marked B. Otherwise z_i takes the value 0.

The transformed variable is $r_i = \frac{z_i-(1-p)\alpha}{p}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Horvitz model. The transformed variable is also reported, if required.

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

HorvitzData

HorvitzDataStCl

HorvitzDataRealSurvey

HorvitzUB

SoberanisCruz

ResamplingVariance

Examples

```
N=10777
data(HorvitzData)
dat=with(HorvitzData,data.frame(z,Pi))
p=0.5
alpha=0.6666667
cl=0.95
Horvitz(dat$z,p,alpha,dat$Pi,"mean",cl,N)

#Horvitz real survey
N=10777
n=710
```

```
data(HorvitzDataRealSurvey)
p=0.5
alpha=1/12
pi=rep(n/N,n)
cl=0.95
Horvitz(HorvitzDataRealSurvey$sex,p,alpha,pi,"mean",cl,N)
```

HorvitzData *Randomized Response Survey on student bullying*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate bullying. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Horvitz model (Horvitz et al., 1967 and Greenberg et al., 1969) with parameter $p = 0.5$. The unrelated question is: Were you born between the 1st and 20th of the month? with $\alpha = 0.6666667$.

Usage

```
HorvitzData
```

Format

A data frame containing a sample of 411 observations from a population of $N = 10777$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Have you been bullied?
- Pi: first-order inclusion probabilities

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

Horvitz

Examples

```
data(HorvitzData)
```

`HorvitzDataRealSurvey` *Randomized Response Survey on a sensitive questions*

Description

This data set contains observations from a randomized response survey conducted in a university to sensitive questions described below. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Horvitz model (Horvitz et al., 1967 and Greenberg et al., 1969) with parameter $p = 0.5$. Each sensitive question is associated with a unrelated question.

1. Were you born in July? with $\alpha = 1/12$
2. Does your ID number end in 2? with $\alpha = 1/10$
3. Were you born of 1 to 20 of the month? with $\alpha = 20/30$
4. Does your ID number end in 5? with $\alpha = 1/10$
5. Were you born of 15 to 25 of the month? with $\alpha = 10/30$
6. Were you born in April? with $\alpha = 1/12$

Usage

```
HorvitzData
```

Format

A data frame containing a sample of 710 observations from a population of $N = 10777$ students. The variables are:

- copied: The randomized response to the question: Have you ever copied in an exam?
- fought: The randomized response to the question: Have you ever fought with a teacher?
- bullied: The randomized response to the question: Have you been bullied?
- bullying: The randomized response to the question: Have you ever bullied someone?
- drug: The randomized response to the question: Have you ever taken drugs on the campus?
- sex: The randomized response to the question: Have you had sex on the premises of the university?

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

`Horvitz`

Examples

```
data(HorvitzDataRealSurvey)
```

`HorvitzDataStCl` *Randomized Response Survey on infidelity*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate the infidelity. The sample is drawn by stratified (by faculty) cluster (by group) sampling. The randomized response technique used is the Horvitz model (Horvitz et al., 1967 and Greenberg et al., 1969) with parameter $p = 0.6$. The unrelated question is: Does your identity card end in an odd number? with a probability $\alpha = 0.5$.

Usage

```
HorvitzDataStCl
```

Format

A data frame containing 365 observations from a population of $N = 1500$ students divided into two strata. The first strata has 14 cluster and the second has 11 cluster. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- CL: Cluster ID
- z: The randomized response to the question: Have you ever been unfaithful?
- Pi: first-order inclusion probabilities

References

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

Horvitz

Examples

```
data(HorvitzDataStCl)
```

`HorvitzUB` *Horvitz-UB model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Horvitz model (Horvitz et al., 1967, and Greenberg et al., 1969) when the proportion of people bearing the innocuous attribute is unknown. The function can also return the transformed variable. The Horvitz-UB model can be seen in Chaudhuri (2011, page 42).

Usage

```
HorvitzUB(I,J,p1,p2,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`I`	first vector of the observed variable; its length is equal to n (the sample size)
`J`	second vector of the observed variable; its length is equal to n (the sample size)
`p1`	proportion of marked cards with the sensitive attribute in the first box
`p2`	proportion of marked cards with the sensitive attribute in the second box
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Horvitz model, when the population proportion α is not known, two independent samples are taken. Two boxes are filled with a large number of similar cards except that in the first box a proportion $p_1(0 < p_1 < 1)$ of them is marked A and the complementary proportion $(1 - p_1)$ each bearing the mark B, while in the second box these proportions are p_2 and $1 - p_2$, maintaining p_2 different from p_1. A sample is chosen and every person sampled is requested to draw one card randomly from the first box and to repeat this independently with the second box. In the first case, a randomized response should be given, as

$$I_i = \begin{cases} 1 & \text{if card type draws "matches" the sensitive trait } A \text{ or the innocuous trait } B \\ 0 & \text{if there is "no match" with the first box} \end{cases}$$

and the second case given a randomized response as

$$J_i = \begin{cases} 1 & \text{if there is "match" for the second box} \\ 0 & \text{if there is "no match" for the second box} \end{cases}$$

The transformed variable is $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Horvitz-UB model. The transformed variable is also reported, if required.

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

HorvitzUBData

Horvitz

ResamplingVariance

Examples

```
N=802
data(HorvitzUBData)
dat=with(HorvitzUBData,data.frame(I,J,Pi))
p1=0.6
p2=0.7
cl=0.95
HorvitzUB(dat$I,dat$J,p1,p2,dat$Pi,"mean",cl,N)
```

HorvitzUBData *Randomized Response Survey on drugs use*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate drugs use. The sample is drawn by cluster sampling with the probabilities proportional to the size. The randomized response technique used is the Horvitz-UB model (Chaudhuri, 2011) with parameters $p_1 = 0.6$ and $p_2 = 0.7$.

Usage

```
HorvitzUBData
```

Format

A data frame containing a sample of 188 observations from a population of $N = 802$ students divided into four cluster. The variables are:

- ID: Survey ID of student respondent
- CL: Cluster ID
- I: The first randomized response to the question: Have you ever used drugs?
- J: The second randomized response to the question: Have you ever used drugs?
- Pi: first-order inclusion probabilities

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Greenberg, B.G., Abul-Ela, A.L., Simmons, W.R., Horvitz, D.G. (1969). *The unrelated question RR model: Theoretical framework.* Journal of the American Statistical Association, 64, 520-539.

Horvitz, D.G., Shah, B.V., Simmons, W.R. (1967). *The unrelated question RR model.* Proceedings of the Social Statistics Section of the American Statistical Association. 65-72. Alexandria, VA: ASA.

See Also

HorvitzUB

Examples

```
data(HorvitzUBData)
```

Kuk *Kuk model*

Description

Computes the randomized response estimation, its variance estimation and its confidence through the Kuk model. The function can also return the transformed variable. The Kuk model was proposed by Kuk in 1990.

Usage

```
Kuk(z,p1,p2,k,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p1	proportion of red cards in the first box
p2	proportion of red cards in the second box
k	total number of cards drawn
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Kuk randomized response technique, the sampled person i is offered two boxes. Each box contains cards that are identical exception colour, either red or white, in sufficiently large numbers with proportions p_1 and $1-p_1$ in the first and p_2 and $1-p_2$, in the second ($p_1 \neq p_2$). The person sampled is requested to use the first box, if his/her trait is A and the second box if his/her trait is A^c and to make k independent draws of cards, with replacement each time. The person is asked to reports $z_i = f_i$, the number of times a red card is drawn.

The transformed variable is $r_i = \frac{f_i/k-p_2}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = br_i + c$, where $b = \frac{1-p_1-p_2}{k(p_1-p_2)}$ and $c = \frac{p_2(1-p_2)}{k(p_1-p_2)^2}$.

Value

Point and confidence estimates of the sensitive characteristics using the Kuk model. The transformed variable is also reported, if required.

References

Kuk, A.Y.C. (1990). *Asking sensitive questions indirectly.* Biometrika, 77, 436-438.

See Also

KukData

ResamplingVariance

Examples

```
N=802
data(KukData)
dat=with(KukData,data.frame(z,Pi))
p1=0.6
p2=0.2
k=25
cl=0.95
Kuk(dat$z,p1,p2,k,dat$Pi,"mean",cl,N)
```

KukData	*Randomized Response Survey on excessive sexual activity*

Description

This data set contains the data from a randomized response survey conducted in a university to investigate excessive sexual activity. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Kuk model (Kuk, 1990) with parameters $p_1 = 0.6$, $p_2 = 0.2$ and $k = 25$.

Usage

```
KukData
```

Format

A data frame containing 200 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you practice excessive sexual activity?
- Pi: first-order inclusion probabilities

References

Kuk, A.Y.C. (1990). *Asking sensitive questions indirectly.* Biometrika, 77, 436-438.

See Also

Kuk

Examples

```
data(KukData)
```

Mangat	*Mangat model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat model. The function can also return the transformed variable. The Mangat model was proposed by Mangat in 1992.

Usage

```
Mangat(z,p,alpha,t,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	proportion of marked cards with the sensitive attribute in the second box
`alpha`	proportion of people with the innocuous attribute
`t`	proportion of marked cards with "True" in the first box
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In Mangat's method, there are two boxes, the first containing cards marked "True" and "RR" in proportions t and $(1-t),(0<t<1)$. A person drawing a "True" marked card is asked to tell the truth about bearing A or A^c. A person drawing and "RR" marked card is then asked to apply Horvitz's device by drawing a card from a second box with cards marked A and B in proportions p and $(1-p)$. If an A marked card is now drawn the truthful response will be about bearing the sensitive attribute A and otherwise about B. The true proportion of people bearing A is to be estimated but α, the proportion of people bearing the innocuous trait B unrelated to A, is assumed to be known. The observed variable is

$$z_i = \begin{cases} y_i & \text{if a card marked "True" is drawn from the first box} \\ I_i & \text{if a card marked "RR" is drawn} \end{cases}$$

where

$$I_i = \begin{cases} 1 & \text{if the type of card drawn from the second box matches trait } A \text{ or } B \\ 0 & \text{if the type of card drawn from the second box does not match trait } A \text{ or } B. \end{cases}$$

The transformed variable is $r_i = \frac{z_i-(1-t)(1-p)\alpha}{t+(1-t)p}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i-1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat model. The transformed variable is also reported, if required.

References

Mangat, N.S. (1992). *Two stage randomized response sampling procedure using unrelated question.* Journal of the Indian Society of Agricultural Statistics, 44, 82-87.

See Also

MangatUB

ResamplingVariance

MangatSingh *Mangat-Singh model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat-Singh model. The function can also return the transformed variable. The Mangat-Singh model was proposed by Mangat and Singh in 1990.

Usage

```
MangatSingh(z,p,t,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive attribute in the second box
t	proportion of marked cards with "True" in the first box
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Mangat-Singh model, the sampled person is offered two boxes of cards. In the first box a known proportion $t, (0 < t < 1)$ of cards is marked "True" and the remaining ones are marked "RR". One card is to be drawn, observed and returned to the box. If the card drawn is marked "True", then the respondent should respond "Yes" if he/she belongs to the sensitive category, otherwise "No". If the card drawn is marked "RR", then the respondent must use the second box and draw a card from it. This second box contains a proportion $p, (0 < p < 1, p \neq 0.5)$ of cards marked A and the remaining ones are marked A^c. If the card drawn from the second box matches his/her status as related to the stigmatizing characteristic, he/she must respond "Yes", otherwise "No". The randomized response from a person labelled i is assumed to be:

$$z_i = \begin{cases} y_i & \text{if a card marked "True" is drawn from the first box} \\ I_i & \text{if a card marked "RR" is drawn} \end{cases}$$

$$I_i = \begin{cases} 1 & \text{if the "card type" } A \text{ or } A^c \text{ "matches" the genuine trait } A \text{ or } A^c \\ 0 & \text{if a "mismatch" is observed} \end{cases}$$

The transformed variable is $r_i = \frac{z_i-(1-t)(1-p)}{t+(1-t)(2p-1)}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-Singh model. The transformed variable is also reported, if required.

References

Mangat, N.S., Singh, R. (1990). *An alternative randomized response procedure.* Biometrika, 77, 439-442.

See Also

MangatSinghData

ResamplingVariance

Examples

```
N=802
data(MangatSinghData)
dat=with(MangatSinghData,data.frame(z,Pi))
p=0.7
t=0.55
cl=0.95
MangatSingh(dat$z,p,t,dat$Pi,"mean",cl,N)
```

MangatSinghData *Randomized Response Survey on cannabis use*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate cannabis use. The sample is drawn by stratified sampling by academic year. The randomized response technique used is the Mangat-Singh model (Mangat and Singh, 1990) with parameters $p = 0.7$ and $t = 0.55$.

Usage

```
MangatSinghData
```

Format

A data frame containing 240 observations from a population of $N = 802$ students divided into four strata. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- z: The randomized response to the question: Have you ever used cannabis?
- Pi: first-order inclusion probabilities

References

Mangat, N.S., Singh, R. (1990). *An alternative randomized response procedure.* Biometrika, 77, 439-442.

See Also

MangatSingh

Examples

```
data(MangatSinghData)
```

MangatSinghSingh *Mangat-Singh-Singh model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat-Singh-Singh model. The function can also return the transformed variable. The Mangat-Singh-Singh model was proposed by Mangat, Singh and Singh in 1992.

Usage

```
MangatSinghSingh(z,p,alpha,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	proportion of marked cards with the sensitive attribute in the box
`alpha`	proportion of people with the innocuous attribute
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In the Mangat-Singh-Singh scheme, a person labelled i, if sampled, is offered a box and told to answer "yes" if the person bears A. But if the person bears A^c then the person is to draw a card from the box with a proportion $p(0 < p < 1)$ of cards marked A and the rest marked B; if the person draws a card marked B he/she is told to say "yes" again if he/she actually bears B; in any other case, "no" is to be answered.

The transformed variable is $r_i = \frac{z_i-(1-p)\alpha}{1-(1-p)\alpha}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i-1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-Singh-Singh model. The transformed variable is also reported, if required.

References

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSinghData

MangatSinghSinghUB

ResamplingVariance

Examples

```
data(MangatSinghSinghData)
dat=with(MangatSinghSinghData,data.frame(z,Pi))
p=0.6
alpha=0.5
cl=0.95
MangatSinghSingh(dat$z,p,alpha,dat$Pi,"total",cl)
```

MangatSinghSinghData *Randomized Response Survey on internet betting*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate internet betting. The sample is drawn by stratified (by faculty) cluster (by group) sampling. The randomized response technique used is the Mangat-Singh-Singh model (Mangat, Singh and Singh, 1992) with parameter $p = 0.6$. The unrelated question is: Does your identity card end in an even number? with a probability $\alpha = 0.5$.

Usage

```
MangatSinghSinghData
```

Format

A data frame containing 802 observations from a population of students divided into eight strata. Each strata has a certain number of clusters, totalling 23. The variables are:

- ID: Survey ID of student respondent
- ST: Strata ID
- CL: Cluster ID

- z: The randomized response to the question: In the last year, did you bet on internet?
- Pi: first-order inclusion probabilities

References

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

`MangatSinghSingh`

Examples

```
data(MangatSinghSinghData)
```

`MangatSinghSinghUB` *Mangat-Singh-Singh-UB model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat-Singh-Singh model (Mangat el al., 1992) when the proportion of people bearing the innocuous attribute is unknown. The function can also return the transformed variable. The Mangat-Singh-Singh-UB model can be seen in Chauduri (2011, page 54).

Usage

```
MangatSinghSinghUB(I,J,p1,p2,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`I`	first vector of the observed variable; its length is equal to n (the sample size)
`J`	second vector of the observed variable; its length is equal to n (the sample size)
`p1`	proportion of marked cards with the sensitive attribute in the first box
`p2`	proportion of marked cards with the sensitive attribute in the second box
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

A person labelled i who is chosen, is instructed to say "yes" if he/she bears A, and if not, to randomly take a card from a box containing cards marked A, B in proportions p_1 and $(1-p_1), (0 < p_1 < 1)$; they are then told to report the value x_i if a B-type card is chosen and he/she bears B; otherwise he/she is told to report "No". This entire exercise is to be repeated independently with the second box with A and B-marked cards in proportions p_2 and $(1-p_2), (0 < p_2 < 1, p_2 \neq p_1)$. Let I_i the first response and J_i the second response for the respondent i.

The transformed variable is $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-Singh-Singh-UB model. The transformed variable is also reported, if required.

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSinghUBData

MangatSinghSingh

ResamplingVariance

Examples

```
N=802
data(MangatSinghSinghUBData)
dat=with(MangatSinghSinghUBData,data.frame(I,J,Pi))
p1=0.6
p2=0.8
cl=0.95
MangatSinghSinghUB(dat$I,dat$J,p1,p2,dat$Pi,"mean",cl,N)
```

MangatSinghSinghUBData *Randomized Response Survey on overuse of the internet*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate overuse of the internet. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Mangat-Singh-Singh-UB model (Chaudhuri, 2011) with parameters $p_1 = 0.6$ and $p_2 = 0.8$.

Usage

```
MangatSinghSinghUBData
```

Format

A data frame containing 500 observations. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you spend a lot of time surfing the internet?
- Pi: first-order inclusion probabilities

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Mangat, N.S., Singh, R., Singh, S. (1992). *An improved unrelated question randomized response strategy.* Calcutta Statistical Association Bulletin, 42, 277-281.

See Also

MangatSinghSinghUB

Examples

```
data(MangatSinghSinghUBData)
```

MangatUB *Mangat-UB model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Mangat model (Mangat, 1992) when the proportion of people bearing the innocuous attribute is unknown. The function can also return the transformed variable. The Mangat-UB model can be seen in Chaudhuri (2011, page 53).

Usage

```
MangatUB(I,J,p1,p2,t,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

I	first vector of the observed variable; its length is equal to n (the sample size)
J	second vector of the observed variable; its length is equal to n (the sample size)
p1	proportion of marked cards with the sensitive attribute in the second box
p2	proportion of marked cards with the sensitive attribute in the third box
t	probability of response to the sensitive questions without using random response in the first box
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

In Mangat's extended scheme, three boxes containing cards are presented to the sampled person, labelled i. The first box contains cards marked "True" and "RR" in proportions t and $1-t$, the second one contains A and B-marked cards in proportions p_1 and $(1-p_1),(0<p_1<1)$ and the third box contains A and B-marked cards in proportions p_2 and $1-p_2,(0<p_2<1), p_1 \neq p_2$. The subject is requested to draw a card from the first box. The sample respondent i is then instructed to tell the truth, using "the first box and if necessary also the second box" and next, independently, to give a second truthful response also using "the first box and if necessary, the third box." Let I_i represent the first response and J_i the second response for respondent i.

The transformed variable is $r_i = \frac{(1-p_2)I_i-(1-p_1)J_i}{p_1-p_2}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Mangat-UB model. The transformed variable is also reported, if required.

References

Chaudhuri, A. (2011). *Randomized response and indirect questioning techniques in surveys.* Boca Raton: Chapman and Hall, CRC Press.

Mangat, N.S. (1992). *Two stage randomized response sampling procedure using unrelated question.* Journal of the Indian Society of Agricultural Statistics, 44, 82-87.

See Also

Mangat

ResamplingVariance

ResamplingVariance *Resampling variance of randomized response models*

Description

To estimate the variance of the randomized response estimators using resampling methods.

Usage

```
ResamplingVariance(output,pi,type=c("total","mean"),option=1,N=NULL,pij=NULL,str=NULL,
clu=NULL,srswr=FALSE)
```

Arguments

`output`	output of the qualitative or quantitative method depending on the variable of interest
`pi`	vector of the first-order inclusion probabilities. By default it is NULL
`type`	the estimator type: total or mean
`option`	method used to calculate the variance (1: Jackknife, 2: Escobar-Berger, 3: Campbell-Berger-Skinner). By default it is 1
`N`	size of the population
`pij`	matrix of the second-order inclusion probabilities. This matrix is necessary for the Escobar-Berger and Campbell-Berger-Skinner options. By default it is NULL
`str`	strata ID. This vector is necessary for the Jackknife option. By default it is NULL
`clu`	cluster ID. This vector is necessary for the Jackknife option. By default it is NULL
`srswr`	variable indicating whether sampling is with replacement. By default it is NULL

Details

Functions to estimate the variance under stratified, cluster and unequal probability sampling by resampling methods (Wolter, 2007). The function ResamplingVariance allows us to choose from three models:

- The Jackknife method (Quenouille, 1949)

- The Escobar-Berger method (Escobar and Berger, 2013)

- The Campbell-Berger-Skinner method (Campbell, 1980; Berger and Skinner, 2005).

The Escobar-Berger and Campbell-Berger-Skinner methods are implemented using the functions defined in samplingVarEst package:

VE.EB.SYG.Total.Hajek, VE.EB.SYG.Mean.Hajek;

VE.Jk.CBS.SYG.Total.Hajek, VE.Jk.CBS.SYG.Mean.Hajek (see López, E., Barrios, E., 2014, for a detailed description of its use).

Note: Both methods require the matrix of the second-order inclusion probabilities. When this matrix is not an input, the program will give a warning and, by default, a jackknife method is used.

Value

The resampling variance of the randomized response technique

References

Berger, Y.G., Skinner, C.J. (2005). *A jackknife variance estimator for unequal probability sampling.* Journal of the Royal Statistical Society B, 67, 79-89.

Campbell, C. (1980). *A different view of finite population estimation.* Proceedings of the Survey Research Methods Section of the American Statistical Association, 319-324.

Escobar, E.L., Berger, Y.G. (2013). *A new replicate variance estimator for unequal probability sampling without replacement.* Canadian Journal of Statistics 41, 3, 508-524.

López, E., Barrios, E. (2014). *samplingVarEst: Sampling Variance Estimation.* R package version 0.9-9. Online http://cran.r-project.org/web/packages/survey/index.html

Quenouille, M.H. (1949). *Problems in Plane Sampling.* The Annals of Mathematical Statistics 20, 355-375.

Wolter, K.M. (2007). *Introduction to Variance Estimation.* 2nd Edition. Springer.

See Also

Warner

ChaudhuriChristofides

EichhornHayre

SoberanisCruz

Horvitz

Examples

```
N=417
data(ChaudhuriChristofidesData)
dat=with(ChaudhuriChristofidesData,data.frame(z,Pi))
mu=c(6,6)
sigma=sqrt(c(10,10))
cl=0.95
data(ChaudhuriChristofidesDatapij)
out=ChaudhuriChristofides(dat$z,mu,sigma,dat$Pi,"mean",cl,pij=ChaudhuriChristofidesDatapij)
out
ResamplingVariance(out,dat$Pi,"mean",2,N,ChaudhuriChristofidesDatapij)

#Resampling with strata
data(EichhornHayreData)
dat=with(EichhornHayreData,data.frame(ST,z,Pi))
mu=1.111111
```

```
sigma=0.5414886
cl=0.95
out=EichhornHayre(dat$z,mu,sigma,dat$Pi,"mean",cl)
out
ResamplingVariance(out,dat$Pi,"mean",1,str=dat$ST)

#Resampling with cluster
N=1500
data(SoberanisCruzData)
dat=with(SoberanisCruzData, data.frame(CL,z,Pi))
p=0.7
alpha=0.5
cl=0.90
out=SoberanisCruz(dat$z,p,alpha,dat$Pi,"total",cl)
out
ResamplingVariance(out,dat$Pi,"total",2,N,samplingVarEst::Pkl.Hajek.s(dat$Pi))

#Resampling with strata and cluster
N=1500
data(HorvitzDataStCl)
dat=with(HorvitzDataStCl, data.frame(ST,CL,z,Pi))
p=0.6
alpha=0.5
cl=0.95
out=Horvitz(dat$z,p,alpha,dat$Pi,"mean",cl,N)
out
ResamplingVariance(out,dat$Pi,"mean",3,N,samplingVarEst::Pkl.Hajek.s(dat$Pi))
```

Saha *Saha model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Saha model. The function can also return the transformed variable. The Saha model was proposed by Saha in 2007.

Usage

```
Saha(z,mu,sigma,pi,type=c("total","mean"),cl,N=NULL,method="srswr")
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
mu	vector with the means of the scramble variables W and U
sigma	vector with the standard deviations of the scramble variables W and U
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean

`cl`	confidence level
`N`	size of the population. By default it is NULL
`method`	method used to draw the sample: srswr or srswor. By default it is srswr

Details

In the Saha model, each respondent selected is asked to report the randomized response $z_i = W(y_i + U)$ where W, U are scramble variables whose distribution is assumed to be known.

To estimate $\bar{Y}$ a sample of respondents is selected according to simple random sampling with replacement. The transformed variable is

$$r_i = \frac{z_i - \mu_W \mu_U}{\mu_W}$$

where μ_W, μ_U are the means of W, U scramble variables respectively

The estimated variance in this model is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n\mu_W^2}$$

where $s_z^2 = \sum_{i=1}^n \frac{(z_i - \bar{z})^2}{n-1}$.

If the sample is selected by simple random sampling without replacement, the estimated variance is

$$\widehat{V}(\widehat{\bar{Y}}_R) = \frac{s_z^2}{n\mu_W^2}\left(1 - \frac{n}{N}\right)$$

Value

Point and confidence estimates of the sensitive characteristics using the Saha model. The transformed variable is also reported, if required.

References

Saha, A. (2007). *A simple randomized response technique in complex surveys.* Metron LXV, 59-66.

See Also

`SahaData`

`ResamplingVariance`

Examples

```
N=228
data(SahaData)
dat=with(SahaData,data.frame(z,Pi))
mu=c(1.5,5.5)
sigma=sqrt(c(1/12,81/12))
cl=0.95
Saha(dat$z,mu,sigma,dat$Pi,"mean",cl,N)
```

SahaData *Randomized Response Survey on spending on alcohol*

Description

This data set contains observations from a randomized response survey conducted in a population of students to investigate spending on alcohol. The sample is drawn by simple random sampling with replacement. The randomized response technique used is the Saha model (Saha, 2007) with scramble variables $W = U(1, 2)$ and $U = U(1, 10)$.

Usage

```
SahaData
```

Format

A data frame containing 100 observations. The variables are:

- ID: Survey ID
- z: The randomized response to the queston: How much money did you spend on alcohol, last weekend?
- Pi: first-order inclusion probabilities

References

Saha, A. (2007). *A simple randomized response technique in complex surveys.* Metron LXV, 59-66.

See Also

Saha

Examples

```
data(SahaData)
```

SinghJoarder *Singh-Joarder model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Singh-Joarder model. The function can also return the transformed variable. The Singh-Joarder model was proposed by Singh and Joarder in 1997.

Usage

```
SinghJoarder(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive question
pi	vector of the first-order inclusion probabilities
type	the estimator type: total or mean
cl	confidence level
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The basics of the Singh-Joarder scheme are similar to Warner's randomized response device, with the following difference. If a person labelled i bears A^c he/she is told to say so if so guided by a card drawn from a box of A and A^c marked cards in proportions p and $(1-p),(0<p<1)$. However, if he/she bears A and is directed by the card to admit it, he/she is told to postpone the reporting based on the first draw of the card from the box but to report on the basis of a second draw. Therefore,

$$z_i = \begin{cases} 1 & \text{if person } i \text{ responds "Yes"} \\ 0 & \text{if person } i \text{ responds "No"} \end{cases}$$

The transformed variable is $r_i = \frac{z_i-(1-p)}{(2p-1)+p(1-p)}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i-1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Singh-Joarder model. The transformed variable is also reported, if required.

References

Singh, S., Joarder, A.H. (1997). *Unknown repeated trials in randomized response sampling.* Journal of the Indian Statistical Association, 30, 109-122.

See Also

SinghJoarderData

ResamplingVariance

Examples

```
N=802
data(SinghJoarderData)
dat=with(SinghJoarderData,data.frame(z,Pi))
p=0.6
cl=0.95
SinghJoarder(dat$z,p,dat$Pi,"mean",cl,N)
```

`SinghJoarderData` *Randomized Response Survey on compulsive spending*

Description

This data set contains observations from a randomized response survey conducted in a university to investigate compulsive spending. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Singh-Joarder model (Singh and Joarder, 1997) with parameter $p = 0.6$.

Usage

```
SinghJoarderData
```

Format

A data frame containing 170 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: Do you have spend compulsively?
- Pi: first-order inclusion probabilities

References

Singh, S., Joarder, A.H. (1997). *Unknown repeated trials in randomized response sampling.* Journal of the Indian Statistical Association, 30, 109-122.

See Also

`SinghJoarder`

Examples

```
data(SinghJoarderData)
```

SoberanisCruz *SoberanisCruz model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the SoberanisCruz model. The function can also return the transformed variable. The SoberanisCruz model was proposed by Soberanis Cruz et al. in 2008.

Usage

```
SoberanisCruz(z,p,alpha,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

z	vector of the observed variable; its length is equal to n (the sample size)
p	proportion of marked cards with the sensitive question
alpha	proportion of people with the innocuous attribute
pi	vector of the first-order inclusion probabilites
type	the estimator type: total or mean
cl	confidence leve
N	size of the population. By default it is NULL
pij	matrix of the second-order inclusion probabilities. By default it is NULL

Details

The SoberanisCruz model considers the introduction of an innocuous variable correlated with the sensitive variable. This variable does not affect individual sensitivity, and maintains reliability. The sampling procedure is the same as in the Horvitz model.

Value

Point and confidence estimates of the sensitive characteristics using the SoberanisCruz model. The transformed variable is also reported, if required.

References

Soberanis Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., González Cossio, F. (2008). *Muestreo de respuestas aleatorizadas en poblaciones finitas: Un enfoque unificador.* Agrociencia Vol. 42 Núm. 5 537-549.

See Also

SoberanisCruzData

Horvitz

ResamplingVariance

Examples

```
data(SoberanisCruzData)
dat=with(SoberanisCruzData,data.frame(z,Pi))
p=0.7
alpha=0.5
cl=0.90
SoberanisCruz(dat$z,p,alpha,dat$Pi,"total",cl)
```

SoberanisCruzData   *Randomized Response Survey on speeding*

Description

This data set contains observations from a randomized response survey conducted in a population of 1500 families in a Spanish town to investigate speeding. The sample is drawn by cluster sampling by district. The randomized response technique used is the SoberanisCruz model (Soberanis Cruz et al., 2008) with parameter $p = 0.7$. The innocuous question is: Is your car medium/high quality? with $\alpha = 0.5$.

Usage

```
SoberanisCruzData
```

Format

A data frame containing 290 observations from a population of $N = 1500$ families divided into twenty cluster. The variables are:

- ID: Survey ID
- CL: Cluster ID
- z: The randomized response to the question: Do you often break the speed limit?
- Pi: first-order inclusion probabilities

References

Soberanis Cruz, V., Ramírez Valverde, G., Pérez Elizalde, S., González Cossio, F. (2008). *Muestreo de respuestas aleatorizadas en poblaciones finitas: Un enfoque unificador.* Agrociencia Vol. 42 Núm. 5 537-549.

See Also

SoberanisCruz

Examples

```
data(SoberanisCruzData)
```

`Warner` *Warner model*

Description

Computes the randomized response estimation, its variance estimation and its confidence interval through the Warner model. The function can also return the transformed variable. The Warner model was proposed by Warner in 1965.

Usage

```
Warner(z,p,pi,type=c("total","mean"),cl,N=NULL,pij=NULL)
```

Arguments

`z`	vector of the observed variable; its length is equal to n (the sample size)
`p`	proportion of marked cards with the sensitive attribute
`pi`	vector of the first-order inclusion probabilities
`type`	the estimator type: total or mean
`cl`	confidence level
`N`	size of the population. By default it is NULL
`pij`	matrix of the second-order inclusion probabilities. By default it is NULL

Details

Warner's randomized response device works as follows. A sampled person labelled i is offered a box of a considerable number of identical cards with a proportion $p, (0 < p < 1, p \neq 0.5)$ of them marked A and the rest marked A^c. The person is requested, randomly, to draw one of them, to observe the mark on the card, and to give the response

$$z_i = \begin{cases} 1 & \text{if card type "matches" the trait } A \text{ or } A^c \\ 0 & \text{if a "no match" results} \end{cases}$$

The randomized response is given by $r_i = \frac{z_i-(1-p)}{2p-1}$ and the estimated variance is $\widehat{V}_R(r_i) = r_i(r_i - 1)$.

Value

Point and confidence estimates of the sensitive characteristics using the Warner model. The transformed variable is also reported, if required.

References

Warner, S.L. (1965). *Randomized Response: a survey technique for eliminating evasive answer bias.* Journal of the American Statistical Association 60, 63-69.

See Also

WarnerData

ResamplingVariance

Examples

```
N=802
data(WarnerData)
dat=with(WarnerData,data.frame(z,Pi))
p=0.7
cl=0.95
Warner(dat$z,p,dat$Pi,"total",cl)
```

WarnerData *Randomized Response Survey on alcohol abuse*

Description

This data set contains observations from a randomized response survey related to alcohol abuse. The sample is drawn by simple random sampling without replacement. The randomized response technique used is the Warner model (Warner, 1965) with parameter $p = 0.7$.

Usage

```
WarnerData
```

Format

A data frame containing 125 observations from a population of $N = 802$ students. The variables are:

- ID: Survey ID of student respondent
- z: The randomized response to the question: During the last month, did you ever have more than five drinks (beer/wine) in succession?
- Pi: first-order inclusion probabilities

References

Warner, S.L. (1965). *Randomized Response: a survey technique for eliminating evasive answer bias.* Journal of the American Statistical Association 60, 63-69.

See Also

Warner

Examples

```
data(WarnerData)
```

Index