

Package ‘vardpoor’

February 14, 2019

Type Package

Title Variance Estimation for Sample Surveys by the Ultimate Cluster Method

Version 0.14.0

Date 2019-02-14

Depends R (>= 3.2.3), data.table (>= 1.10.4), laeken, stringr, ggplot2, pracma, graphics

Imports foreach, plyr, MASS, stats, utils, surveyplanning

Maintainer Juris Breidaks <Juris.Breidaks@csb.gov.lv>

Description Generation of domain variables, linearization of several nonlinear population statistics (the ratio of two totals, weighted income percentile, relative median income ratio, at-risk-of-poverty rate, at-risk-of-poverty threshold, Gini coefficient, gender pay gap, the aggregate replacement ratio, the relative median income ratio, median income below at-risk-of-poverty gap, income quintile share ratio, relative median at-risk-of-poverty gap), computation of regression residuals in case of weight calibration, variance estimation of sample surveys by the ultimate cluster method (Hansen, Hurwitz and Madow, Theory, vol. I: Methods and Applications; vol. II: Theory. 1953, New York: John Wiley and Sons), variance estimation for longitudinal, cross-sectional measures and measures of change for single and multi-stage cluster sampling designs (Berger, Y. G., 2015, <doi:10.1111/rssa.12116>). Several other precision measures are derived - standard error, the coefficient of variation, the margin of error, confidence interval, design effect.

License GPL (>= 2)

Encoding latin1

Repository CRAN

URL <https://csblatvia.github.io/vardpoor/>

BugReports <https://github.com/CSBLatvia/vardpoor/issues/>

NeedsCompilation yes

Author Juris Breidaks [aut, cre],
Martins Liberts [aut],
Santa Ivanova [aut]

Date/Publication 2019-02-14 12:10:02 UTC

R topics documented:

vardpoor-package	2
domain	4
incPercentile	6
lin.ratio	7
linarpr	8
linarpt	10
linarr	13
lingini	15
lingini2	17
lingpg	19
linpoormed	21
linqsr	23
linrmir	25
linrmpg	27
residual_est	29
vardannual	31
vardchanges	38
vardchangespoor	43
vardchangstrs	48
vardcros	50
vardcrossoor	57
vardom	62
vardomh	66
vardom_othstr	71
variance_est	74
variance_othstr	77
varpoord	79
var_srs	85
Index	87

vardpoor-package	<i>Variance Estimation for Sample Surveys by the Ultimate Cluster Method</i>
------------------	--

Description

Generation of domain variables, linearization of several nonlinear population statistics (the ratio of two totals, weighted income percentile, at risk of poverty rate, at risk of poverty threshold, Gini coefficient, gender pay gap, the aggregate replacement ratio, the relative median income ratio, median income of individuals below at risk of poverty gap, income quintile share ratio, relative median at risk of poverty gap), computation of regression residuals in case of weight calibration, variance estimation of sample surveys by the ultimate cluster method (Hansen, Hurwitz and Madow, Theory, vol. I: Methods and Applications; vol. II: Theory. 1953, New York: John Wiley and Sons.), variance estimation for longitudinal, cross - sectional measures and measures of change for single and multistage stage cluster sampling designs (Berger, Y. G. ,2015, <doi.org/10.1111/rssa.12116>).

Several other precision measures are derived - standard error, the coefficient of variation, the margin of error, confidence interval, design effect.

Details

Package: vardpoor
 Version: 0.14.0
 Date: 2019-02-14
 Depends: R (>= 3.2.3), data.table (>= 1.10.4), laeken, stringr
 Imports: reshape, foreach, plyr, MASS, stats, utils, surveyplanning, ggplot2, pracma, graphics
 License: GPL (>= 2)
 URL: <https://github.com/CSBLatvia/vardpoor/>
 BugReports: <https://github.com/CSBLatvia/vardpoor/issues/>

Index:

domain	Split data variables by domains
incPercentile	Weighted income percentile
lin.ratio	Linearization of the ratio estimator
linarr	Linearization of the aggregate replacement ratio
linrmi	Linearization of the relative median income ratio
linarpr	Linearization of the at-risk-of-poverty rate
linarpt	Linearization of the at-risk-of-poverty threshold
lingini	Linearization of the GINI coefficient I
lingini2	Linearization of the GINI coefficient II
lingpg	Linearization of the gender pay (wage) gap
linpoormed	Linearization of the median income below the at-risk-of-poverty gap
linqsr	Linearization of the income quintile share ratio
linrmpg	Linearization of the relative median at-risk-of-poverty gap
residual_est	Residual estimation of calibration
var_srs	The estimation of the simple random sampling
vardchanges	Variance estimation for measures of change for single and multistage stage cluster sampling designs
vardchangespoor	Variance estimation for measures of change for sample surveys for indicators on social exclusion and poverty
vardchangstrs	Variance estimation for measures of change for stratified simple random sampling
vardcros	Variance estimation for cross-sectional and longitudinal measures for single and multistage stage cluster sampling designs
vardcrossoor	Variance estimation for cross-sectional, longitudinal measures for indicators on social exclusion and poverty
vardom	Estimation of the variance in domain by the ultimate cluster method

vandom_othstr	Variance estimation for sample surveys in domain using two stratification
vandomh	Variance estimation for sample surveys in domain for one or two stage surveys by the ultimate cluster method
vardpoor-package	Variance estimation for sample surveys by the ultimate cluster method <code>~~ package title ~~</code>
variance_est	Variance estimation for sample surveys by the ultimate cluster method
variance_othstr	Variance estimation for sample surveys by the new stratification
varpoord	Estimation of the variance and deff for sample surveys for indicators on social exclusion and poverty

Author(s)

Juris Breidaks [aut, cre], Martins Liberts [aut] Santa Ivanova [aut]

Maintainer: Juris Breidaks <Juris.Breidaks@csb.gov.lv>

domain	<i>Extra variables for domain estimation</i>
--------	--

Description

The function computes extra variables for domain estimation. Each unique D row defines a domain. Extra variables are computed for each Y variable.

Usage

```
domain(Y, D, dataset = NULL, checking = TRUE)
```

Arguments

Y	Matrix of study variables. Any object convertible to <code>data.table</code> with numeric values, NA values are not allowed. Object convertible to <code>data.table</code> or variable names as character, column numbers.
D	Matrix of domain variables. Any object convertible to <code>data.table</code> . The number of rows of D must match the number of rows of Y. Duplicated names are not allowed. Object convertible to <code>data.table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to <code>data.table</code> .
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

Numeric `data.table` containing extra variables for domain estimation.

References

Carl-Erik Sarndal, Bengt Swensson, Jan Wretman. Model Assisted Survey Sampling. Springer-Verlag, 1992, p.70.

See Also

[vandom](#), [vandomh](#)

Examples

```
### Example 0
```

```
domain(Y = 1, D = "A")
```

```
### Example 1
```

```
Y1 <- as.matrix(1 : 10)
colnames(Y1) <- "Y1"
D1 <- as.matrix(rep(1, 10))
colnames(D1) <- "D1"
domain(Y = Y1, D = D1)
```

```
### Example 2
```

```
Y <- matrix(1 : 20, 10, 2)
colnames(Y) <- paste0("Y", 1 : 2)
D <- matrix(rep(1 : 2, each = 5), 10, 1)
colnames(D) <- "D"
domain(Y, D)
```

```
### Example 3
```

```
Y <- matrix(1 : 20, 10, 2)
colnames(Y) <- paste0("Y", 1 : 2)
D <- matrix(rep(1 : 4, each = 5), 10, 2)
colnames(D) <- paste0("D", 1 : 2)
domain(Y, D)
```

```
### Example 4
```

```
Y <- matrix(1 : 20, 10, 2)
colnames(Y) <- paste0("Y", 1 : 2)
D <- matrix(c(rep(1 : 2, each = 5), rep(3, 10)), 10, 2)
colnames(D) <- paste0("D", 1 : 2)
domain(Y, D)
```

incPercentile	<i>Estimation of weighted percentiles</i>
---------------	---

Description

The function computes the estimates of weighted percentiles.

Usage

```
incPercentile(Y, weights = NULL, sort = NULL,
             Dom = NULL, period = NULL,
             k = c(20, 80), dataset = NULL,
             checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
weights	Optional weight variable. One dimensional object convert to one-column <code>data.table</code> or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, the estimates of percentiles are computed for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, linearization of at-risk-of-poverty threshold is done for each survey period. Object convertible to <code>data.table</code> or variable names as character, column numbers as numeric vector.
k	A vector of values between 0 and 100 specifying the percentiles to be computed (0 gives the minimum, 100 gives the maximum).
dataset	Optional survey data object convertible to <code>data.table</code> .
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A `data.table` containing the estimates of weighted income percentiles specified by `k`.

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

See Also

[linarpt](#), [linarpr](#), [linqsr](#)

Examples

```
data(eusilc)
incPercentile(Y = "eqIncome", weights = "rb050", Dom = "db040", dataset = eusilc)
```

lin.ratio

Linearization of the ratio estimator

Description

Computes linearized variable for the ratio estimator.

Usage

```
lin.ratio(Y, Z, weight, Dom = NULL, dataset = NULL, percentratio = 1, checking = TRUE)
```

Arguments

Y	Matrix of numerator variables. Any object convertible to <code>data.table</code> with numeric values, NA values are not allowed.
Z	Matrix of denominator variables. Any object convertible to <code>data.table</code> with numeric values, NA values are not allowed.
weight	Weight variable. One dimensional object convertible to one-column <code>data.table</code> .
Dom	Optional variables used to define population domains. If supplied, the linearized variables are computed for each domain. An object convertible to <code>data.table</code> .
dataset	Optional survey data object convertible to <code>data.table</code> .
percentratio	Positive integer value. All linearized variables are multiplied with <code>percentratio</code> value, by default - 1.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

The function returns the `data.table` of the linearized variables for the ratio estimator.

References

Carl-Erik Sarndal, Bengt Swensson, Jan Wretman. Model Assisted Survey Sampling. Springer-Verlag, 1992, p.178.

See Also

[domain](#), [vardom](#), [vardomh](#), [vardcros](#), [vardchanges](#), [vardannual](#)

Examples

```

Y <- data.table(Y = rchisq(10, 3))
Z <- data.table(Z = rchisq(10, 3))
weights <- rep(2, 10)
data.table(Y, Z, weights,
           V1 = lin.ratio(Y, Z, weights, percentratio = 1),
           V10 = lin.ratio(Y, Z, weights, percentratio = 10),
           V100 = lin.ratio(Y, Z, weights, percentratio = 100))

```

linarpr

Linearization of at-risk-of-poverty rate

Description

Estimates the at-risk-of-poverty rate (defined as the proportion of persons with equalized disposable income below at-risk-of-poverty threshold) and computes linearized variable for variance estimation.

Usage

```

linarpr(Y, id = NULL, weight = NULL,
        Y_thres = NULL, wght_thres = NULL,
        sort = NULL, Dom = NULL, period = NULL,
        dataset = NULL, percentage = 60,
        order_quant = 50, var_name = "lin_arpr",
        kern_method = "gaussian", r = NULL, ro = NULL,
        h_breaks = NULL, checking = TRUE)

```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column data.table or variable name as character, column number).
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data.table or variable name as character, column number or logical vector).
weight	Optional weight variable. One dimensional object convertible to one-column data.table or variable name as character, column number or logical vector).
Y_thres	Variable (for example equalized disposable income) used for computation and linearization of poverty threshold. One dimensional object convertible to one-column data.table or variable name as character, column number. Variable specified for inc is used as income_thres if income_thres is not defined.
wght_thres	Weight variable used for computation and linearization of poverty threshold. One dimensional object convertible to one-column data.table or variable name as character, column number or logical vector. Variable specified for weight is used as wght_thres if wght_thres is not defined.

sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data.table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of at-risk-of-poverty threshold is done for each domain. An object convertible to data.table or variable names as character vector, column numbers as numeric vector.
period	Optional variable for survey period. If supplied, linearization of at-risk-of-poverty threshold is done for each survey period. Object convertible to data.table or variable names as character, column numbers as numeric vector.
dataset	Optional survey data object convertible to data.table.
percentage	A numeric value in range [0, 100] for p in the formula for at-risk-of-poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ <p>For example, to compute at-risk-of-poverty threshold equal to 60% of some income quantile, p should be set equal to 60.</p>
order_quant	A numeric value in range [0, 100] for α in the formula for at-risk-of-poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ <p>For example, to compute at-risk-of-poverty threshold equal to some percentage of median income, α should be set equal to 50.</p>
var_name	A character specifying the name of the linearized variable.
kern_method	A character specifying the method used for density calculation.
r	Value specifying the smoothing spline degree.
ro	Value specifying the smoothing spline error from histogram.
h_breaks	Value specifying the histogram bins.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

The implementation strictly follows the Eurostat definition.

Value

A list with four objects are returned:

quantile	A data.table containing the estimated value of the quintile used for at-risk-of-poverty threshold estimation.
threshold	A data.table containing the estimated at-risk-of-poverty threshold.
value	A data.table containing the estimated at-risk-of-poverty rate (in percentage).
lin	A data.table containing the linearized variables of the at-risk-of-poverty rate (in percentage).

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[linarpt](#), [varpoord](#), [vardcros poor](#), [vardchanges poor](#)

Examples

```
data(eusilc)

dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
d <- linarpr(Y = "eqIncome", id = "IDd",
            weight = "rb050", Dom = NULL,
            dataset = dati, percentage = 60,
            order_quant = 50L)

d$value

## Not run:
# By domains
dd <- linarpr(Y = "eqIncome", id = "IDd",
             weight = "rb050", Dom = "db040",
             dataset = dati, percentage = 60,
             order_quant=50L)

## End(Not run)
```

linarpt

Linearization of at-risk-of-poverty threshold

Description

Estimates the at-risk-of-poverty threshold (defined as percentage (usually 60%) of equalised disposable income after social transfers quantile (usually median)) and computes linearized variable for variance estimation.

Usage

```
linarpt(Y, id = NULL, weight = NULL,
        sort = NULL, Dom = NULL, period = NULL,
        dataset = NULL, percentage = 60,
        order_quant = 50, var_name = "lin_arpt",
        kern_method = "gaussian", r = NULL,
        ro = NULL, h_breaks = NULL,
        checking = TRUE)
```

Arguments

Y	Study variable (for example equalised disposable income after social transfers). One dimensional object convertible to one-column data . table or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data . table or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data . table or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data . table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of at-risk-of-poverty threshold is done for each domain. An object convertible to data . table or variable names as character vector, column numbers as numeric vector.
period	Optional variable for survey period. If supplied, linearization of at-risk-of-poverty threshold is done for each survey period. Object convertible to data . table or variable names as character, column numbers as numeric vector.
dataset	Optional survey data object convertible to data . table.
percentage	A numeric value in range $[0, 100]$ for p in the formula for at-risk-of-poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to 60% of some income quantile, p should be set equal to 60.
order_quant	A numeric value in range $[0, 100]$ for α in the formula for at-risk-of-poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to some percentage of median income, α should be set equal to 50.
var_name	A character specifying the name of the linearized variable.
kern_method	A character specifying the method used for density calculation.
r	Value specifying the smoothing spline degree.
ro	Value specifying the smoothing spline error from histogram.

h_breaks	Value specifying the histogram bins.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

The implementation strictly follows the Eurostat definition.

Value

A list with three objects are returned:

quantile	A data . table containing the estimated value of the quintile used for at-risk-of-poverty threshold estimation.
value	A data . table containing the estimated at-risk-of-poverty threshold (in percentage).
lin	A data . table containing the linearized variables of the at-risk-of-poverty threshold (in percentage).

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[linarpr](#), [incPercentile](#), [varpoord](#), [vardcros poor](#), [vardchanges poor](#)

Examples

```
data(eusilc)

dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
d1 <- linarpt(Y = "eqIncome", id = "IDd",
             weight = "rb050", Dom = NULL,
             dataset = dati, percentage = 60,
             order_quant = 50L)
```

```

d1$value

## Not run:
# By domains
d2 <- linarrpt(Y = "eqIncome", id = "IDd",
              weight = "rb050", Dom = "db040",
              dataset = dati, percentage = 60,
              order_quant = 50L)

d2$value

## End(Not run)

```

linarr

Linearization of the aggregate replacement ratio

Description

Estimates the aggregate replacement ratio (defined as the gross median individual pension income of the population aged 65-74 relative to the gross median individual earnings from work of the population aged 50-59, excluding other social benefits) and computes linearized variable for variance estimation.

Usage

```

linarr(Y, Y_den, id = NULL, age, pl085,
       month_at_work, weight = NULL,
       sort = NULL, Dom = NULL, period = NULL,
       dataset = NULL, order_quant = 50,
       var_name = "lin_arr", checking = TRUE)

```

Arguments

Y	Numerator variable (for gross pension income). One dimensional object convertible to one-column data.table or variable name as character, column number.
Y_den	Denominator variable (for example gross individual earnings). One dimensional object convertible to one-column data.table or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data.table or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data.table or variable name as character, column number.
age	Age variable. One dimensional object convertible to one-column data.table or variable name as character, column number.

p1085	Retirement variable (Number of months spent in retirement or early retirement). One dimensional object convertible to one-column data . table or variable name as character, column number.
month_at_work	Variable for total number of month at work (sum of the number of months spent at full-time work as employee, number of months spent at part-time work as employee, number of months spent at full-time work as self-employed (including family worker), number of months spent at part-time work as self-employed (including family worker)). One dimensional object convertible to one-column data . table or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data . table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of at-risk-of-poverty threshold is done for each domain. An object convertible to data . table or variable names as character vector, column numbers as numeric vector.
period	Optional variable for survey period. If supplied, linearization of at-risk-of-poverty threshold is done for each survey period. Object convertible to data . table or variable names as character, column numbers as numeric vector.
dataset	Optional survey data object convertible to data . table.
order_quant	A numeric value in range [0, 100] for α in the formula for at-risk-of-poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute at-risk-of-poverty threshold equal to some percentage of median income, α should be set equal to 50.
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

The implementation strictly follows the Eurostat definition.

Value

A list with four objects are returned:

value	A data . table containing the estimated the aggregate replacement ratio.
lin	A data . table containing the linearized variables of the aggregate replacement ratio.

References

Working group on Statistics on Income and Living Conditions (2015) Task 5 - Improvement and optimization of calculation of net change. *LC- 139/15/EN*, Eurostat.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[varpoord](#), [vardcrospoor](#), [vardchangespoor](#)

Examples

```
data(eusilc)

dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)
dati$pl085 <- 12 * trunc(runif(nrow(dati), 0, 2))
dati$month_at_work <- 12 * trunc(runif(nrow(dati), 0, 2))

# Full population
d <- linarr(Y = "eqIncome", Y_den = "eqIncome",
           id = "IDd", age = "age",
           pl085 = "pl085", month_at_work = "month_at_work",
           weight = "rb050", Dom = NULL,
           dataset = dati, order_quant = 50L)

d$value

## Not run:
# By domains
dd <- linarr(Y = "eqIncome", Y_den = "eqIncome",
            id = "IDd", age = "age",
            pl085 = "pl085", month_at_work = "month_at_work",
            weight = "rb050", Dom = "db040",
            dataset = dati, order_quant = 50L)

dd

## End(Not run)
```

lingini

Linearization of the GINI coefficient I

Description

Estimate the Gini coefficient, which is a measure for inequality, and its linearization.

Usage

```
lingini(Y, id = NULL, weight = NULL,
       sort = NULL, Dom = NULL, period = NULL,
       dataset = NULL, var_name = "lin_gini",
       checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column data.table or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data.table or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data.table or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data.table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of the GINI is done for each domain. An object convertible to data.table or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, linearization of the GINI is done for each time period. Object convertible to data.table or variable names as character, column numbers.
dataset	Optional survey data object convertible to data.table.
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with two objects are returned by the function:

value	A data.table containing the estimated Gini coefficients (in percentage) by G. Osier and Eurostat.
lin	A data.table containing the linearized variables of the Gini coefficients (in percentage) by G. Osier.

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[lingini2](#), [linqsr](#), [varpoord](#), [vardcros poor](#), [vardchanges poor](#)

Examples

```
data(eusilc)
dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)[1 : 3,]

# Full population
dat1 <- lingini(Y = "eqIncome", id = "IDd", weight = "rb050", dataset = dati)
dat1$value

## Not run:
# By domains
dat2 <- lingini(Y = "eqIncome", id = "IDd", weight = "rb050", Dom = c("db040"), dataset = dati)
dat2$value
## End(Not run)
```

lingini2

Linearization of the GINI coefficient II

Description

Estimate the Gini coefficient, which is a measure for inequality, and its linearization.

Usage

```
lingini2(Y, id = NULL, weight = NULL,
         sort = NULL, Dom = NULL,
         period = NULL, dataset = NULL,
         var_name = "lin_gini2",
         checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

Dom	Optional variables used to define population domains. If supplied, linearization of the GINI is done for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, linearization of the GINI is done for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to <code>data.table</code> .
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with two objects are returned by the function:

value	A <code>data.table</code> containing the estimated Gini coefficients (in percentage) by Langel and Tille (2012) and Eurostat.
lin	A <code>data.table</code> containing the linearized variables of the Gini coefficients (in percentage) by Langel and Tille (2012).

References

Eric Graf and Yves Tille, Variance Estimation Using Linearization for Poverty and Social Exclusion Indicators, *Survey Methodology*, June 2014 61 Vol. 40, No. 1, pp. 61-79, Statistics Canada, Catalogue no. 12-001-X, URL <http://www.statcan.gc.ca/pub/12-001-x/12-001-x2014001-eng.pdf>

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

MATTI LANGEL - YVES TILLE, Corrado Gini, a pioneer in balanced sampling and inequality theory. *METRON - International Journal of Statistics*, 2011, vol. LXIX, n. 1, pp. 45-65, URL <ftp://metron.sta.uniroma1.it/RePEc/articoli/2011-1-3.pdf>.

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

See Also

[lingini](#), [linqsr](#), [varpoord](#), [vardcros poor](#), [vardchanges poor](#)

Examples

```

data(eusilc)
dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
dat1 <- lingini2(Y = "eqIncome", id = "IDd", weight = "rb050", dataset = dati)
dat1$value

## Not run:
# By domains
dat2 <- lingini2(Y = "eqIncome", id = "IDd", weight = "rb050", Dom = c("db040"), dataset = dati)
dat2$value
## End(Not run)

```

lingpg

Linearization of the gender pay (wage) gap.

Description

Estimation of gender pay (wage) gap and computation of linearized variables for variance estimation.

Usage

```

lingpg(Y, gender = NULL, id = NULL,
       weight = NULL, sort = NULL, Dom = NULL,
       period = NULL, dataset = NULL,
       var_name = "lin_gpg",
       checking = TRUE)

```

Arguments

Y	Study variable (for example the gross hourly earning). One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
gender	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

Dom	Optional variables used to define population domains. If supplied, estimation and linearization of gender pay (wage) gap is done for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, estimation and linearization of gender pay (wage) gap is done for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to <code>data.table</code> .
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with two objects are returned:

value	<code>data.table</code> containing the estimated gender pay (wage) gap (in percentage).
lin	<code>data.table</code> containing the linearized variables of the gender pay (wage) gap (in percentage) for variance estimation.

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[lingqr](#), [lingini](#), [varpoord](#), [vardcrospoor](#), [vardchangespoor](#)

Examples

```
data(ses)
```

```
dati <- data.table(ID = paste0("V", 1 : nrow(ses)), ses)
setnames(dati, "sex", "sexf")
dati[sexf == "male", sex:= 1]
dati[sexf == "female", sex:= 2]
```

```

# Full population
gpgs1 <- lingpg(Y = "earningsHour", gender = "sex",
               id = "ID", weight = "weights",
               dataset = dati)

gpgs1$value

## Not run:
# Domains by education
gpgs2 <- lingpg(Y = "earningsHour", gender = "sex",
               id = "ID", weight = "weights",
               Dom = "education", dataset = dati)

gpgs2$value

# Sort variable
gpgs3 <- lingpg(Y = "earningsHour", gender = "sex",
               id = "ID", weight = "weights",
               sort = "ID", Dom = "education",
               dataset = dati)

gpgs3$value

# Two survey periods
dati[, year := 2010]
dati2 <- copy(dati)
dati2[, year := 2011]
dati <- rbind(dati, dati2)
gpgs4 <- lingpg(Y = "earningsHour", gender = "sex",
               id = "ID", weight = "weights",
               sort = "ID", Dom = "education",
               period = "year", dataset = dati)

gpgs4$value
names(gpgs4$lin)

## End(Not run)

```

linpoormed

Linearization of the median income of individuals below the At Risk of Poverty Threshold

Description

Estimation of the median income of individuals below At Risk of Poverty Threshold and computation of linearized variable for variance estimation. The At Risk of Poverty Threshold is estimated for the whole population always. The median income is estimated for the whole population or for each domain.

Usage

```

linpoormed(Y, id = NULL, weight = NULL,
           sort = NULL, Dom = NULL, period = NULL,

```

```
dataset = NULL, percentage = 60,
order_quant = 50, var_name = "lin_poormed",
checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column data.table or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data.table or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data.table or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data.table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of the median income of persons below a poverty threshold is done for each domain. An object convertible to data.table or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, linearization of the median income of persons below a poverty threshold is done for each time period. Object convertible to data.table or variable names as character, column numbers.
dataset	Optional survey data object convertible to data.table.
percentage	A numeric value in range [0,100] for p in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to 60% of some income quantile, p should be set equal to 60.
order_quant	A numeric value in range [0,100] for α in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to some percentage of median income, α should be set equal to 50.
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with two objects are returned by the function:

value	A data.table containing the estimated median income of individuals below the At Risk of Poverty Threshold.
lin	A data.table containing the linearized variables of the median income below the At Risk of Poverty Threshold.

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[linarpt](#), [linrmpg](#), [varpoord](#), [vardcrospoor](#), [vardchangespoor](#)

Examples

```
data(eusilc)
dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
d <- linpoormed(Y = "eqIncome", id = "IDd",
               weight = "rb050", Dom = NULL,
               dataset = dati, percentage = 60,
               order_quant = 50L)

## Not run:
# Domains by location of household
dd <- linpoormed(Y = "eqIncome", id = "IDd",
                 weight = "rb050", Dom = "db040",
                 dataset = dati, percentage = 60,
                 order_quant = 50L)

## End(Not run)
```

Description

Estimate the Quintile Share Ratio, which is defined as the ratio of the sum of equalized disposable income received by the top 20% to the sum of equalized disposable income received by the bottom 20%, and its linearization.

Usage

```
linqsr(Y, id = NULL, weight = NULL,
       sort = NULL, Dom = NULL, period = NULL,
       dataset = NULL, alpha = 20,
       var_name = "lin_qsr", checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column data. table or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data. table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of the income quintile share ratio is done for each domain. An object convertible to data. table or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, linearization of the income quintile share ratio is done for each time period. Object convertible to data. table or variable names as character, column numbers.
dataset	Optional survey data object convertible to data. table.
alpha	a numeric value in range $[0, 100]$ for the order of the Quintile Share Ratio.
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with two objects are returned by the function:

value	A data. table containing the estimated Quintile Share Ratio by G. Osier and Eurostat papers.
lin	A data. table containing the linearized variables of the Quintile Share Ratio by G. Osier paper.

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361,

URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[incPercentile](#), [varpoord](#), [vardcros poor](#), [vardchanges poor](#)

Examples

```
data(eusilc)
dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
dd <- linqsr(Y = "eqIncome", id = "IDd", weight = "rb050",
            Dom = NULL, dataset = dati, alpha = 20)
dd$value

## Not run:
# By domains
dd <- linqsr(Y = "eqIncome", id = "IDd", weight = "rb050",
            Dom = "db040", dataset = dati, alpha = 20)
dd$value
## End(Not run)
```

linrmir

Linearization of the relative median income ratio

Description

Estimates the relative median income ratio (defined as the ratio of the median equivalised disposable income of people aged above age to the median equivalised disposable income of those aged below 65) and computes linearized variable for variance estimation.

Usage

```
linrmir(Y, id = NULL, age,
        weight = NULL,
        sort = NULL,
        Dom = NULL,
        period = NULL,
        dataset = NULL,
        order_quant = 50,
        var_name = "lin_rmir",
        checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
age	Age variable. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of at-risk-of-poverty threshold is done for each domain. An object convertible to data. <code>table</code> or variable names as character vector, column numbers as numeric vector.
period	Optional variable for survey period. If supplied, linearization of at-risk-of-poverty threshold is done for each survey period. Object convertible to data. <code>table</code> or variable names as character, column numbers as numeric vector.
dataset	Optional survey data object convertible to data. <code>table</code> .
order_quant	A numeric value in range $[0, 100]$ for α in the formula for at-risk-of-poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ <p>For example, to compute the relative median income ratio to some percentage of median income, α should be set equal to 50.</p>
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

The implementation strictly follows the Eurostat definition.

Value

A list with four objects are returned:

value	A data. <code>table</code> containing the estimated relative median income ratio.
lin	A data. <code>table</code> containing the linearized variables of the relative median income ratio.

References

Working group on Statistics on Income and Living Conditions (2015) Task 5 - Improvement and optimization of calculation of net change. *LC- 139/15/EN*, Eurostat.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[varpoord](#), [vardcrospoor](#), [vardchangespoor](#)

Examples

```
data(eusilc)

dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
d <- linrmir(Y = "eqIncome", id = "IDd", age = "age",
            weight = "rb050", Dom = NULL,
            dataset = dati, order_quant = 50L)
d$value

## Not run:
# By domains
dd <- linrmir(Y = "eqIncome", id = "IDd", age = "age",
            weight = "rb050", Dom = "db040",
            dataset = dati, order_quant = 50L)
dd
## End(Not run)
```

 linrmpg

Linearization of the relative median at-risk-of-poverty gap

Description

Estimate the relative median at-risk-of-poverty gap, which is defined as the relative difference between the median equalized disposable income of persons below the At Risk of Poverty Threshold and the At Risk of Poverty Threshold itself (expressed as a percentage of the at-risk-of-poverty threshold) and its linearization.

Usage

```
linrmpg(Y, id = NULL, weight = NULL,
        sort = NULL, Dom = NULL, period = NULL,
        dataset = NULL, percentage = 60,
```

```
order_quant = 50, var_name = "lin_rmpg",
checking = TRUE)
```

Arguments

Y	Study variable (for example equalized disposable income). One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
weight	Optional weight variable. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data. <code>table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of the relative median at-risk-of-poverty gap is done for each domain. An object convertible to data. <code>table</code> or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, linearization of the relative median at-risk-of-poverty gap is done for each time period. Object convertible to data. <code>table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to data. <code>table</code> .
percentage	A numeric value in range [0,100] for p in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to 60% of some income quantile, p should be set equal to 60.
order_quant	A numeric value in range [0,100] for α in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to some percentage of median income, α should be set equal to 50.
var_name	A character specifying the name of the linearized variable.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with two objects are returned by the function:

value	A data. <code>table</code> containing the estimated relative median at-risk-of-poverty gap (in percentage).
lin	A data. <code>table</code> containing the linearized variables of the relative median at-risk-of-poverty gap (in percentage).

References

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[linarpt](#), [linarpr](#), [linpoormed](#), [varpoord](#), [vardcros poor](#), [vardchanges poor](#)

Examples

```
data(eusilc)
dati <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

# Full population
d <- linrmpg(Y = "eqIncome", id = "IDd",
            weight = "rb050", Dom = NULL,
            dataset = dati, percentage = 60,
            order_quant = 50L)

d$value
d$threshold

## Not run:
# By domains
dd <- linrmpg(Y = "eqIncome", id = "IDd",
             weight = "rb050", Dom = "db040",
             dataset = dati, percentage = 60,
             order_quant = 50L)

dd$value
## End(Not run)
```

 residual_est

Residual estimation of calibration

Description

Computes the estimation residuals of calibration.

Usage

```
residual_est(Y, X, weight, q, dataset = NULL, checking = TRUE)
```

Arguments

Y	Matrix of the variable of interest.
X	Matrix of the auxiliary variables for the calibration estimator. This is the matrix of the sample calibration variables.
weight	Weight variable. One dimensional object convertible to one-column data.frame.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column data.frame.
dataset	Optional survey data object convertible to data.table.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

The function implements the following estimator:

$$e_k = Y_k - X_k' B$$

where

$$\hat{B} = \left(\sum_s weight_k q_k X_k X_k' \right)^{-1} \left(\sum_s weight_k q_k X_k Y_k \right)$$

Value

A list with objects are returned by the function:

residuals	A numeric data.table containing the estimated residuals of calibration.
betas	A numeric data.table containing the estimated coefficients of calibration.

References

Sixten Lundstrom and Carl-Erik Sarndal. Estimation in the presence of Nonresponse and Frame Imperfections. Statistics Sweden, 2001, p. 43-44.

See Also

[domain](#), [lin.ratio](#), [linarpr](#), [linarpt](#), [lingini](#), [lingini2](#), [lingpg](#), [linpoormed](#), [linqsr](#), [linmpg](#), [vardom](#), [vardomh](#), [varpoord](#), [variance_est](#), [variance_othstr](#)

Examples

```

Y <- matrix(rchisq(10, 3), 10, 1)
X <- matrix(rchisq(20, 3), 10, 2)
w <- rep(2, 10)
q <- rep(1, 10)
residual_est(Y, X, w, q)

### Test2
Y <- matrix(rchisq(10, 3), 10, 1)
X <- matrix(c(rchisq(10, 2), rchisq(10, 2) + 10), 10, 2)
w <- rep(2, 10)
q <- rep(1, 10)
residual_est(Y, X, w, q)
as.matrix(lm(Y ~ X - 1, weights = w * q)$residuals)

```

vardannual	<i>Variance estimation for measures of annual net change or annual for single and multistage stage cluster sampling designs</i>
------------	---

Description

Computes the variance estimation for measures of annual net change or annual for single and multistage stage cluster sampling designs.

Usage

```

vardannual(Y, H, PSU, w_final, ID_level1,
           ID_level2, Dom = NULL, Z = NULL,
           gender = NULL, country = NULL,
           years, subperiods, dataset = NULL,
           year1 = NULL, year2 = NULL, X = NULL,
           countryX = NULL, yearsX = NULL,
           subperiodsX = NULL, X_ID_level1 = NULL,
           ind_gr = NULL, g = NULL, q = NULL,
           datasetX = NULL, frate = 0, percentratio = 1,
           use.estVar = FALSE, use.gender = FALSE,
           confidence = 0.95, method = "cros")

```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

w_final	Weight variable. One dimensional object convertible to one-column data . table or variable name as character, column number.
ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column data . table or variable name as character, column number.
ID_level2	Optional variable for unit ID codes. One dimensional object convertible to one-column data . table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables are calculated for each domain. An object convertible to data . table or variable names as character vector, column numbers.
Z	Optional variables of denominator for ratio estimation. If supplied, the ratio estimation is computed. Object convertible to data . table or variable names as character, column numbers. This variable is NULL by default.
gender	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column data . table or variable name as character, column number.
country	Variable for the survey countries. The values for each country are computed independently. Object convertible to data . table or variable names as character, column numbers.
years	Variable for the all survey years. The values for each year are computed independently. Object convertible to data . table or variable names as character, column numbers.
subperiods	Variable for the all survey subperiods. The values for each subperiod are computed independently. Object convertible to data . table or variable names as character, column numbers.
dataset	Optional survey data object convertible to data . table.
year1	The vector of years from variable years describes the first year for measures of annual net change.
year2	The vector of years from variable periods describes the second year for measures of annual net change.
X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to data . table or variable names as character, column numbers.
countryX	Optional variable for the survey countries. The values for each country are computed independently. Object convertible to data . table or variable names as character, column numbers.
yearsX	Variable of the all survey years. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to data . table or variable names as character, column numbers.
subperiodsX	Variable for the all survey subperiods. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to data . table or variable names as character, column numbers.
X_ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column data . table or variable name as character, column number.

<code>ind_gr</code>	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column data. table or variable name as character, column number.
<code>g</code>	Optional variable of the g weights. One dimensional object convertible to one-column data. table or variable name as character, column number.
<code>q</code>	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column data. table or variable name as character, column number.
<code>datasetX</code>	Optional survey data object in household level convertible to data. table.
<code>frate</code>	Positive numeric value. Sampling rate in percentage, by default - 0.
<code>percentratio</code>	Positive numeric value. All linearized variables are multiplied with percentratio value, by default - 1.
<code>use.estVar</code>	Logical value. If value is TRUE, then R function <code>estVar</code> is used for the estimation of covariance matrix of the residuals. If value is FALSE, then R function <code>estVar</code> is not used for the estimation of covariance matrix of the residuals.
<code>use.gender</code>	Logical value. If value is TRUE, then subperiods is defined together with gender.
<code>confidence</code>	optional; either a positive value for confidence interval. This variable by default is 0.95.
<code>method</code>	character value; value 'cros' is for measures of annual or value 'netchanges' is for measures of annual net change. This variable by default is netchanges.

Value

A list with objects are returned by the function:

`crosssectional_results`

A data. table containing: `year` - survey years,
`subperiods` - survey subperiods,
`country` - survey countries,
`Dom` - optional variable of the population domains,
`namesY` - variable with names of variables of interest,
`namesZ` - optional variable with names of denominator for ratio estimation,
`sample_size` - the sample size (in numbers of individuals),
`pop_size` - the population size (in numbers of individuals),
`total` - the estimated totals,
`variance` - the estimated variance of cross-sectional or longitudinal measures,
`sd_w` - the estimated weighted variance of simple random sample,
`sd_nw` - the estimated variance estimation of simple random sample,
`pop` - the population size (in numbers of households),
`sampl_siz` - the sample size (in numbers of households),
`stderr_w` - the estimated weighted standard error of simple random sample,
`stderr_nw` - the estimated standard error of simple random sample,
`se` - the estimated standard error of cross-sectional or longitudinal,
`rse` - the estimated relative standard error (coefficient of variation),
`cv` - the estimated relative standard error (coefficient of variation) in percentage,
`absolute_margin_of_error` - the estimated absolute margin of error,

relative_margin_of_error - the estimated relative margin of error,
 CI_lower - the estimated confidence interval lower bound,
 CI_upper - the estimated confidence interval upper bound,
 confidence_level - the positive value for confidence interval.

crosssectional_var_grad

A data.table containing year - survey years,
 subperiods - survey subperiods,
 country - survey countries,
 Dom - optional variable of the population domains,
 namesY - variable with names of variables of interest,
 namesZ - optional variable with names of denominator for ratio estimation,
 grad - the estimated gradient,
 var - the estimated a design-based variance.

vardchanges_grad_var

A data.table containing year_1 - survey years of years1,
 subperiods_1 - survey subperiods of years1,
 year_2 - survey years of years2,
 subperiods_2 - survey subperiods of years2,
 country - survey countries,
 Dom - optional variable of the population domains,
 namesY - variable with names of variables of interest,
 namesZ - optional variable with names of denominator for ratio estimation,
 nams - gradient names, numenator (num) and denominator (den), for each year,
 grad - the estimated gradient,
 cros_var - the estimated a design-based variance.

vardchanges_rho

A data.table containing year - survey years of years for crosssectional estimates,
 subperiods - survey subperiods of years for crosssectional estimates,
 year_1 - survey years of years1,
 subperiods_1 - survey subperiods of years1,
 year_2 - survey years of years2,
 subperiods_2 - survey subperiods of years2,
 country - survey countries,
 Dom - optional variable of the population domains,
 namesY - variable with names of variables of interest,
 namesZ - optional variable with names of denominator for ratio estimation,
 nams - gradient names, numenator (num) and denominator (den), for each year,
 rho - the estimated correlation matrix.

vardchanges_var_tau

A data.table containing year_1 - survey years of years1,
 subperiods_1 - survey subperiods of years1,
 year_2 - survey years of years2,
 subperiods_2 - survey subperiods of years2,
 country - survey countries,
 Dom - optional variable of the population domains,
 namesY - variable with names of variables of interest,
 namesZ - optional variable with names of denominator for ratio estimation,

	<p>nams - gradient names, numenator (num) and denominator (den), for each year, var_tau - the estimated covariance matrix.</p>
vardchanges_results	<p>A data.table containing year - survey years of years for measures of annual, subperiods - survey subperiods of years for measures of annual, year_1 - survey years of years1 for measures of annual net change, subperiods_1 - survey subperiods of years1 for measures of annual net change, year_2 - survey years of years2 for measures of annual net change, subperiods_2 - survey subperiods of years2 for measures of annual net change, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, estim_1 - the estimated value for period1, estim_2 - the estimated value for period2, estim - the estimated value, var - the estimated variance, se - the estimated standard error, CI_lower - the estimated confidence interval lower bound, CI_upper - the estimated confidence interval upper bound, confidence_level - the positive value for confidence interval, significant - is the the difference significant</p>
X_annual	<p>A data.table containing year - survey years of years for measures of annual, year_1 - survey years of years1 for measures of annual net change, year_2 - survey years of years2 for measures of annual net change, period - period1 and period2 together, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, cros_se - the estimated cross-sectional standard error.</p>
A_matrix	<p>A data.table containing year - survey years of years1 for measures of annual, year_1 - survey years of years1 for measures of annual net change, year_2 - survey years of years2 for measures of annual net change, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, cols - the estimated matrix_A columns, matrix_A - the estimated matrix A.</p>
annual_sum	<p>A data.table containing year - survey years, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, totalY - the estimated value of variables of interest for period1, totalZ - optional the estimated value of denominator for period2, estim - the estimated value for year .</p>

annual_results A data.table containing year - survey years of years for measures of annual,
 year_1 - survey years of years1 for measures of annual net change,
 year_2 - survey years of years2 for measures of annual net change,
 country - survey countries,
 Dom - optional variable of the population domains,
 namesY - variable with names of variables of interest,
 namesZ - optional variable with names of denominator for ratio estimation,
 estim_1 - the estimated value for period1 for measures of annual net change,
 estim_2 - the estimated value for period2 for measures of annual net change,
 estim - the estimated value,
 var - the estimated variance,
 se - the estimated standard error,
 rse - the estimated relative standard error (coefficient of variation),
 cv - the estimated relative standard error (coefficient of variation) in percentage,
 absolute_margin_of_error - the estimated absolute margin of error for period1 for measures of annual,
 relative_margin_of_error - the estimated relative margin of error in percentage for measures of annual,
 CI_lower - the estimated confidence interval lower bound,
 CI_upper - the estimated confidence interval upper bound,
 confidence_level - the positive value for confidence interval,
 significant - is the the difference significant

References

Guillaume OSIER, Virginie RAYMOND, (2015), Development of methodology for the estimate of variance of annual net changes for LFS-based indicators. Deliverable 1 - Short document with derivation of the methodology.

Guillaume Osier, Yves Berger, Tim Goedeme, (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, Eurostat Methodologies and Working papers, URL <http://ec.europa.eu/eurostat/documents/3888793/5855973/KS-RA-13-024-EN.PDF>.

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

See Also

[domain](#), [vardcros](#), [vardchanges](#)

Examples

```

### Example
data("eusilc")
set.seed(1)
eusilc1 <- eusilc[1 : 20,]
set.seed(1)
data <- data.table(rbind(eusilc1, eusilc1),
                  year = c(rep(2010, nrow(eusilc1)),
                          rep(2011, nrow(eusilc1))))

data[, country:= "AT"]
data[, half:= .I - 2 * trunc((.I - 1) / 2)]
data[, quarter:= .I - 4 * trunc((.I - 1) / 4)]
data[age < 0, age:= 0]
PSU <- data[, .N, keyby = "db030"][, N:= NULL]
PSU[, PSU:= trunc(runif(nrow(PSU), 0, 5))]
data <- merge(data, PSU, all = TRUE, by = "db030")
PSU <- eusilc <- NULL
data[, strata := c("XXXX")]

data[, employed := trunc(runif(nrow(data), 0, 2))]
data[, unemployed := trunc(runif(nrow(data), 0, 2))]
data[, labour_force := employed + unemployed]
data[, id_lv2 := paste0("V", .I)]
result <- vardannual(Y = "employed", H = "strata",
                    PSU = "PSU", w_final = "rb050",
                    ID_level1 = "db030", ID_level2 = "id_lv2",
                    Dom = NULL, Z = NULL, years = "year", subperiods = "half",
                    dataset = data, percentratio = 100,
                    confidence = 0.95, method = "cros")

## Not run:

result2 <- vardannual(Y = "employed", H = "strata",
                     PSU = "PSU", w_final = "rb050",
                     ID_level1 = "db030", ID_level2 = "id_lv2",
                     Dom = NULL, Z = NULL, country = "country",
                     years = "year", subperiods = "quarter",
                     dataset = data, year1 = 2010, year2 = 2011,
                     percentratio = 100, confidence = 0.95,
                     method = "netchanges")

result <- vardannual(Y = "unemployed", H = "strata",
                    PSU = "PSU", w_final = "rb050",
                    ID_level1 = "db030", ID_level1 = "id_lv2",
                    Dom = NULL, Z = "labour_force",
                    country = "country", years = "year",
                    subperiods = "quarter", dataset = data,
                    year1 = 2010, year2 = 2011,
                    percentratio = 100, confidence = 0.95,
                    method = "netchanges")

## End(Not run)

```

vardchanges	<i>Variance estimation for measures of change for single and multistage stage cluster sampling designs</i>
-------------	--

Description

Computes the variance estimation for measures of change for single and multistage stage cluster sampling designs.

Usage

```
vardchanges(Y, H, PSU, w_final,
            ID_level1, ID_level2,
            Dom = NULL, Z = NULL,
            gender = NULL, country = NULL,
            period, dataset = NULL,
            period1, period2,
            X = NULL, countryX = NULL,
            periodX = NULL, X_ID_level1 = NULL,
            ind_gr = NULL, g = NULL,
            q = NULL, datasetX = NULL,
            linratio = FALSE,
            percentratio = 1,
            use.estVar = FALSE,
            outp_res = FALSE,
            confidence = 0.95,
            change_type = "absolute",
            checking = TRUE)
```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ID_level2	Optional variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables are calculated for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.

Z	Optional variables of denominator for ratio estimation. If supplied, the ratio estimation is computed. Object convertible to <code>data.table</code> or variable names as character, column numbers. This variable is NULL by default.
gender	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
country	Variable for the survey countries. The values for each country are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
period	Variable for the all survey periods. The values for each period are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to <code>data.table</code> .
period1	The vector of periods from variable <code>periods</code> describes the first period.
period2	The vector of periods from variable <code>periods</code> describes the second period.
X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
countryX	Optional variable for the survey countries. The values for each country are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
periodX	Optional variable of the all survey periods. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
X_ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
datasetX	Optional survey data object in household level convertible to <code>data.table</code> .
linratio	Logical value. If value is TRUE, then the linearized variables for the ratio estimator is used for variance estimation. If value is FALSE, then the gradients is used for variance estimation.
percentratio	Positive numeric value. All linearized variables are multiplied with <code>percentratio</code> value, by default - 1.
use.estVar	Logical value. If value is TRUE, then R function <code>estVar</code> is used for the estimation of covariance matrix of the residuals. If value is FALSE, then R function <code>estVar</code> is not used for the estimation of covariance matrix of the residuals.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.

confidence	optional; either a positive value for confidence interval. This variable by default is 0.95 .
change_type	character value net changes type - absolute or relative.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with objects are returned by the function:

res_out	A data.table containing the estimated residuals of calibration with ID_level1 and PSU by periods and countries (if available).
crosssectional_results	A data.table containing: period - survey periods, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, sample_size - the sample size (in numbers of individuals), pop_size - the population size (in numbers of individuals), total - the estimated totals, variance - the estimated variance of cross-sectional or longitudinal measures, sd_w - the estimated weighted variance of simple random sample, sd_nw - the estimated variance estimation of simple random sample, pop - the population size (in numbers of households), sampl_siz - the sample size (in numbers of households), stderr_w - the estimated weighted standard error of simple random sample, stderr_nw - the estimated standard error of simple random sample, se - the estimated standard error of cross-sectional or longitudinal, rse - the estimated relative standard error (coefficient of variation), cv - the estimated relative standard error (coefficient of variation) in percentage, absolute_margin_of_error - the estimated absolute margin of error, relative_margin_of_error - the estimated relative margin of error, CI_lower - the estimated confidence interval lower bound, CI_upper - the estimated confidence interval upper bound.
crosssectional_var_grad	A data.table containing periods - survey periods, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, grad - the estimated gradient, var - the estimated a design-based variance.
rho	A data.table containing periods_1 - survey periods of periods1, periods_2 - survey periods of periods2, country - survey countries, Dom - optional variable of the population domains,

namesY - variable with names of variables of interest,
namesZ - optional variable with names of denominator for ratio estimation,
nams - the variable names in correlation matrix,
rho - the estimated correlation matrix.

var_tau A data.table containing periods_1 - survey periods of periods1,
periods_2 - survey periods of periods2,
country - survey countries,
Dom - optional variable of the population domains,
namesY - variable with names of variables of interest,
namesZ - optional variable with names of denominator for ratio estimation,
nams - the variable names in correlation matrix,
var_tau - the estimated covariance matrix.

changes_results A data.table containing periods_1 - survey periods of periods1,
periods_2 - survey periods of periods2,
country - survey countries,
Dom - optional variable of the population domains,
namesY - variable with names of variables of interest,
namesZ - optional variable with names of denominator for ratio estimation,
estim_1 - the estimated value for period1,
estim_2 - the estimated value for period2,
estim - the estimated value,
var - the estimated variance,
se - the estimated standard error,
CI_lower - the estimated confidence interval lower bound,
CI_upper - the estimated confidence interval upper bound.
significant - is the the difference significant

References

Guillaume Osier, Yves Berger, Tim Goedeme, (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, Eurostat Methodologies and Working papers, URL <http://ec.europa.eu/eurostat/documents/3888793/5855973/KS-RA-13-024-EN.PDF>.

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

See Also

[domain](#), [vardcros](#), [vardchangespoor](#)

Examples

```

### Example

data("eusilc")
set.seed(1)
eusilc1 <- eusilc[1:40,]

set.seed(1)
data <- data.table(rbind(eusilc1, eusilc1),
                  year = c(rep(2010, nrow(eusilc1)),
                          rep(2011, nrow(eusilc1))))

data[age < 0, age := 0]
PSU <- data[, .N, keyby = "db030"][, N := NULL]
PSU[, PSU := trunc(runif(nrow(PSU), 0, 5))]
data <- merge(data, PSU, all = TRUE, by = "db030")
PSU <- eusilc <- NULL
data[, strata := c("XXXX")]

data[, t_pov := trunc(runif(nrow(data), 0, 2))]
data[, exp := 1]

# At-risk-of-poverty (AROP)
data[, pov := ifelse (t_pov == 1, 1, 0)]
data[, id_lev2 := paste0("V", .I)]

result <- vardchanges(Y = "pov", H = "strata",
                     PSU = "PSU", w_final = "rb050",
                     ID_level1 = "db030", ID_level2 = "id_lev2",
                     Dom = NULL, Z = NULL, period = "year",
                     dataset = data, period1 = 2010,
                     period2 = 2011, change_type = "absolute")

## Not run:
data("eusilc")
data <- data.table(rbind(eusilc, eusilc),
                  year=c(rep(2010, nrow(eusilc)),
                        rep(2011, nrow(eusilc))))

data[age < 0, age := 0]
PSU <- data[, .N, keyby = "db030"][, N := NULL]
PSU[, PSU := trunc(runif(nrow(PSU), 0, 100))]
data <- merge(data, PSU, all = TRUE, by = "db030")
PSU <- eusilc <- NULL
data[, strata := "XXXX"]

data[, t_pov := trunc(runif(nrow(data), 0, 2))]
data[, t_dep := trunc(runif(nrow(data), 0, 2))]
data[, t_lwi := trunc(runif(nrow(data), 0, 2))]
data[, exp := 1]
data[, exp2 := 1 * (age < 60)]

```

```

# At-risk-of-poverty (AROP)
data[, pov := ifelse (t_pov == 1, 1, 0)]

# Severe material deprivation (DEP)
data[, dep := ifelse (t_dep == 1, 1, 0)]

# Low work intensity (LWI)
data[, lwi := ifelse (t_lwi == 1 & exp2 == 1, 1, 0)]

# At-risk-of-poverty or social exclusion (AROPE)
data[, arope := ifelse (pov == 1 | dep == 1 | lwi == 1, 1, 0)]
data[, dom := 1]
data[, id_lev2 := .I]

result <- vardchanges(Y = c("pov", "dep", "lwi", "arope"),
  H = "strata", PSU = "PSU", w_final = "rb050",
  ID_level1 = "db030", ID_level2 = "id_lev2",
  Dom = "rb090", Z = NULL, period = "year",
  dataset = data, period1 = 2010,
  period2 = 2011, change_type = "absolute")

## End(Not run)

```

vardchangespoor

Variance estimation for measures of change for sample surveys for indicators on social exclusion and poverty

Description

Computes the variance estimation for measures of change for indicators on social exclusion and poverty.

Usage

```

vardchangespoor(Y, age = NULL, pl085 = NULL,
  month_at_work = NULL, Y_den = NULL,
  Y_thres = NULL, wght_thres = NULL,
  H, PSU, w_final, ID_level1, ID_level2,
  Dom = NULL, country = NULL, period,
  sort = NULL, period1, period2,
  gender = NULL, dataset = NULL, X = NULL,
  countryX = NULL, periodX = NULL,
  X_ID_level1 = NULL, ind_gr = NULL,
  g = NULL, q = NULL, datasetX = NULL,
  percentage = 60, order_quant = 50,
  alpha = 20, use.estVar = FALSE,
  confidence = 0.95, outp_lin = FALSE,
  outp_res = FALSE, type = "linrmpg",
  change_type = "absolute")

```

Arguments

Y	Study variable (for example equalized disposable income or gross pension income). One dimensional object convertible to one-column data. table or variable name as character, column number.
age	Age variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
p1085	Retirement variable (Number of months spent in retirement or early retirement). One dimensional object convertible to one-column data. table or variable name as character, column number.
month_at_work	Variable for total number of month at work (sum of the number of months spent at full-time work as employee, number of months spent at part-time work as employee, number of months spent at full-time work as self-employed (including family worker), number of months spent at part-time work as self-employed (including family worker)). One dimensional object convertible to one-column data. table or variable name as character, column number.
Y_den	Denominator variable (for example gross individual earnings). One dimensional object convertible to one-column data. table or variable name as character, column number.
Y_thres	Variable (for example equalized disposable income) used for computation and linearization of poverty threshold. One dimensional object convertible to one-column data. table or variable name as character, column number. Variable specified for inc is used as income_thres if income_thres is not defined.
wght_thres	Weight variable used for computation and linearization of poverty threshold. One dimensional object convertible to one-column data. table or variable name as character, column number. Variable specified for weight is used as wght_thres if wght_thres is not defined.
H	The unit stratum variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column data. table or variable name as character, column number or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset).
ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
ID_level2	Optional variable for unit ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables are calculated for each domain. An object convertible to data. table or variable names as character vector, column numbers.
country	Variable for the survey countries. The values for each country are computed independently. Object convertible to data. table or variable names as character, column numbers.

period	Variable for the all survey periods. The values for each period are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
period1	The vector from variable <code>period</code> describes the first period.
period2	The vector from variable <code>period</code> describes the second period.
gender	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
dataset	Optional survey data object convertible to <code>data.frame</code> .
X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
countryX	Optional variable for the survey countries. The values for each country are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
periodX	Optional variable of the survey periods and countries. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
X_ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
datasetX	Optional survey data object in household level convertible to <code>data.table</code> .
percentage	A numeric value in range $[0, 100]$ for p in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to 60% of some income quantile, p should be set equal to 60.
order_quant	A numeric value in range $[0, 100]$ for α in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to some percentage of median income, α should be set equal to 50.

alpha	a numeric value in range [0, 100] for the order of the income quantile share ratio (in percentage).
use.estVar	Logical value. If value is TRUE, then R function estVar is used for the estimation of covariance matrix of the residuals. If value is FALSE, then R function estVar is not used for the estimation of covariance matrix of the residuals.
confidence	optional; either a positive value for confidence interval. This variable by default is 0.95.
outp_lin	Logical value. If TRUE linearized values of the ratio estimator will be printed out.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.
type	a character vector (of length one unless several.ok is TRUE), example "linarpr", "linarpt", "lingpg", "linpoormed", "linrmpg", "lingini", "lingini2", "linqsr", "linarr", "linr-mir", "all_choices".
change_type	character value net changes type - absolute or relative.

Value

A list with objects are returned by the function:

cros_lin_out	A data.table containing the linearized values of the ratio estimator with ID_level2 and PSU by periods and countries (if available).
cros_res_out	A data.table containing the estimated residuals of calibration with ID_level1 and PSU by periods and countries (if available).
crossectional_results	A data.table containing: period - survey periods, country - survey countries, Dom - optional variable of the population domains, type - type variable, count_respondents - the count of respondents, pop_size - the population size (in numbers of individuals), estim - the estimated value, se - the estimated standard error, var - the estimated variance, rse - the estimated relative standard error (coefficient of variation), cv - the estimated relative standard error (coefficient of variation) in percentage.
changes_results	A data.table containing: period - survey periods, country - survey countries, Dom - optional variable of the population domains, type - type variable, estim_1 - the estimated value for period1, estim_2 - the estimated value for period2, estim - the estimated value, se - the estimated standard error,


```

PSU = "PSU", w_final="rb050",
ID_level1 = "db030", ID_level2 = "id_l2",
Dom = c("rb090"), country = "country",
period = "year", sort = NULL,
period1 = c(2010, 2011),
period2 = c(2011, 2010),
gender = NULL, dataset = data,
percentage = 60, order_quant = 50L,
alpha = 20, confidence = 0.95,
type = "linrmpg")

```

vardchangstrs	<i>Variance estimation for measures of change for stratified simple random sampling</i>
---------------	---

Description

Computes the variance estimation for measures of change for stratified simple random sampling.

Usage

```

vardchangstrs(Y, H, PSU, w_final,
              Dom = NULL, periods = NULL,
              dataset, periods1, periods2,
              in_sample, in_frame,
              confidence = 0.95,
              percentratio = 1)

```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables are calculated for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
periods	Variable for the all survey periods. The values for each period are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to <code>data.table</code> .

periods1	The vector of periods from variable periods describes the first period for measures of change.
periods2	The vector of periods from variable periods describes the second period for measures of change.
in_sample	Sample variable. One dimensional object convertible to one-column data . table or variable name as character, column number.
in_frame	Frame variable. One dimensional object convertible to one-column data . table or variable name as character, column number.
percentratio	Positive numeric value. All linearized variables are multiplied with percentratio value, by default - 1.
confidence	optional; either a positive value for confidence interval. This variable by default is 0.95.

Value

A list with objects are returned by the function:

annual_sum	<p>A data . table containing year - survey years, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, totalY - the estimated value of variables of interest for period1, totalZ - optional the estimated value of denominator for period2, year - survey years of years for measures of annual, year_1 - survey years of years1 for measures of annual net change, year_2 - survey years of years2 for measures of annual net change, country - survey countries, Dom - optional variable of the population domains, namesY - variable with names of variables of interest, namesZ - optional variable with names of denominator for ratio estimation, estim_1 - the estimated value for period1 for measures of annual net change, estim_2 - the estimated value for period2 for measures of annual net change, estim - the estimated value, var - the estimated variance, se - the estimated standard error, rse - the estimated relative standard error (coefficient of variation), cv - the estimated relative standard error (coefficient of variation) in percentage, absolute_margin_of_error - the estimated absolute margin of error for period1 for measures of annual, relative_margin_of_error - the estimated relative margin of error in percentage for measures of annual, CI_lower - the estimated confidence interval lower bound, CI_upper - the estimated confidence interval upper bound, confidence_level - the positive value for confidence interval, significant - is the the difference significant</p>
------------	---

References

Nordberg, L. ((2000), On Variance Estimation for Measures of Change When Samples are Coordinated by the Use of Permanent Random Numbers. *Journal of Official Statistics* 16, 363:378

See Also

[domain](#)

Examples

```
### Example
data("eusilc")
set.seed(1)
eusilc1 <- eusilc[1 : 100,]
set.seed(1)
datas <- data.table(rbind(eusilc1, eusilc1),
                    year = c(rep(2010, nrow(eusilc1)),
                            rep(2011, nrow(eusilc1))))

PSU <- datas[, .N, keyby = "db030"][, N:= NULL]
PSU[, PSU:= trunc(runif(nrow(PSU), 0, 5))]
datas <- merge(datas, PSU, all = TRUE, by = "db030")
datas[, strata := c("XXXX")]

datas[, calc := trunc(runif(nrow(datas), 0, 2))]
datas[, id_lv2 := paste0("V", .I)]

## Not run:
result <- vardchangstrs(Y = "calc", H = "strata",
                       PSU = "PSU", w_final = "rb050",
                       ID_level1 = "db030", ID_level2 = "id_lv2",
                       periods = "year", dataset = datas,
                       periods1 = 2010, periods2 = 2011,
                       percentratio = 100, confidence = 0.95)

## End(Not run)
```

vardcros

Variance estimation for cross-sectional, longitudinal measures for single and multistage stage cluster sampling designs

Description

Computes the variance estimation for cross-sectional and longitudinal measures for any stage cluster sampling designs.

Usage

```
vardcros(Y, H, PSU, w_final, ID_level1,
         ID_level2, Dom = NULL, Z = NULL,
         gender = NULL, country = NULL,
         period, dataset = NULL, X = NULL,
         countryX = NULL, periodX = NULL,
         X_ID_level1 = NULL, ind_gr = NULL,
         g = NULL, q = NULL, datasetX = NULL,
         linratio = FALSE, percentratio=1,
         use.estVar = FALSE, ID_level1_max = TRUE,
         outp_res = FALSE, withperiod = TRUE,
         netchanges = TRUE, confidence = .95,
         checking = TRUE)
```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ID_level2	Optional variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables are calculated for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
Z	Optional variables of denominator for ratio estimation. If supplied, the ratio estimation is computed. Object convertible to <code>data.table</code> or variable names as character, column numbers. This variable is NULL by default.
gender	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
country	Variable for the survey countries. The values for each country are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
period	Variable for the survey periods. The values for each period are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
dataset	Optional survey data object convertible to <code>data.table</code> .

X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
countryX	Optional variable for the survey countries. The values for each country are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
periodX	Optional variable of the survey periods and countries. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
X_ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
datasetX	Optional survey data object in household level convertible to <code>data.table</code> .
linratio	Logical value. If value is TRUE, then the linearized variables for the ratio estimator is used for variance estimation. If value is FALSE, then the gradients is used for variance estimation.
percentratio	Positive numeric value. All linearized variables are multiplied with <code>percentratio</code> value, by default - 1.
use.estVar	Logical value. If value is TRUE, then R function <code>estVar</code> is used for the estimation of covariance matrix of the residuals. If value is FALSE, then R function <code>estVar</code> is not used for the estimation of covariance matrix of the residuals.
ID_level1_max	Logical value. If value is TRUE, then the size of sample for variance under simple random sampling is taken as maximum value of size in <code>ID_level1</code> . If value is FALSE, then the size of sample for variance under simple random sampling is taken as count of <code>ID_level2</code> in <code>ID_level1</code> .
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.
withperiod	Logical value. If TRUE is value, the results is with period, if FALSE, without period.
netchanges	Logical value. If value is TRUE, then produce two objects: the first object is aggregation of weighted data by period (if available), country, strata and PSU, the second object is an estimation for Y, the variance, gradient for numerator and denominator by country and period (if available). If value is FALSE, then both objects containing NULL.
confidence	Optional positive value for confidence interval. This variable by default is 0.95.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with four objects are returned by the function:

res_out	A data . table containing the estimated residuals of calibration with ID_level1 and PSU.
data_net_changes	A data . table containing aggregation of weighted data by period (if available) and countries (if available), country, strata, PSU.
var_grad	A data . table containing estimation for Y, the variance, gradient for numerator and denominator by period, country (if available) and population domains (if available).
results	A data . table containing period - survey periods, country - survey countries (if available), Dom - optional variable of the population domains, namesY - names of variables of interest, namesZ - optional variable for names of denominator for ratio estimation, sample_size - the sample size (in numbers of individuals), pop_size - the population size (in numbers of individuals), total - the estimated totals, variance - the estimated variance of cross-sectional or longitudinal measures, sd_w - the estimated weighted variance of simple random sample, sd_nw - the estimated variance estimation of simple random sample, pop - the population size (in numbers of households), sampl_siz - the sample size (in numbers of households), stderr_w - the estimated weighted standard error of simple random sample, stderr_nw - the estimated standard error of simple random sample, se - the estimated standard error of cross-sectional or longitudinal, rse - the estimated relative standard error (coefficient of variation), cv - the estimated relative standard error (coefficient of variation) in percentage, absolute_margin_of_error - the estimated absolute margin of error, relative_margin_of_error - the estimated relative margin of error, CI_lower - the estimated confidence interval lower bound, CI_upper - the estimated confidence interval upper bound, confidence_level - the positive value for confidence interval.

References

Guillaume Osier, Yves Berger, Tim Goedeme, (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, Eurostat Methodologies and Working papers, URL <http://ec.europa.eu/eurostat/documents/3888793/5855973/KS-RA-13-024-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/>

[3859598/5927001/KS-RA-13-029-EN.PDF](#).

See Also

[domain](#), [lin.ratio](#)

Examples

```
# Example 1
data(eusilc)
set.seed(1)
data <- data.table(eusilc)
data[, year := 2010]
data[, country := "AT"]
data[age < 0, age := 0]
PSU <- data[, .N, keyby = "db030"][, N := NULL]
PSU[, PSU := trunc(runif(nrow(PSU), 0, 100))]
data <- merge(data, PSU, by = "db030", all = TRUE)
PSU <- eusilc <- 0

data[, strata := "XXXX"]
data[, t_pov := trunc(runif(nrow(data), 0, 2))]
data[, t_dep := trunc(runif(nrow(data), 0, 2))]
data[, t_lwi := trunc(runif(nrow(data), 0, 2))]
data[, exp := 1]
data[, exp2 := 1 * (age < 60)]

# At-risk-of-poverty (AROP)
data[, pov := ifelse (t_pov == 1, 1, 0)]

# Severe material deprivation (DEP)
data[, dep := ifelse (t_dep == 1, 1, 0)]

# Low work intensity (LWI)
data[, lwi := ifelse (t_lwi == 1 & exp2 == 1, 1, 0)]

# At-risk-of-poverty or social exclusion (AROPE)
data[, arope := ifelse (pov == 1 | dep == 1 | lwi == 1, 1, 0)]

result11 <- vardcros(Y="arope", H = "strata",
                    PSU = "PSU", w_final = "rb050",
                    ID_level1 = "db030", ID_level2 = "rb030",
                    Dom = "rb090", Z = NULL, country = "country",
                    period = "year", dataset = data,
                    linratio = FALSE, withperiod = TRUE,
                    netchanges = TRUE, confidence = .95)

## Not run:
# Example 2
data(eusilc)
set.seed(1)
data <- data.table(rbind(eusilc, eusilc),
```

```

        year = c(rep(2010, nrow(eusilc)),
                 rep(2011, nrow(eusilc)))

data[, country := "AT"]
data[age < 0, age := 0]
PSU <- data[, .N, keyby = "db030"][, N := NULL]
PSU[, PSU := trunc(runif(nrow(PSU), 0, 100))]
data <- merge(data, PSU, by = "db030", all = TRUE)
PSU <- eusilc <- 0
data[, strata := "XXXX"]
data[, strata := as.character(strata)]
data[, t_pov := trunc(runif(nrow(data), 0, 2))]
data[, t_dep := trunc(runif(nrow(data), 0, 2))]
data[, t_lwi := trunc(runif(nrow(data), 0, 2))]
data[, exp := 1]
data[, exp2 := 1 * (age < 60)]

# At-risk-of-poverty (AROP)
data[, pov := ifelse(t_pov == 1, 1, 0)]

# Severe material deprivation (DEP)
data[, dep := ifelse(t_dep == 1, 1, 0)]

# Low work intensity (LWI)
data[, lwi := ifelse(t_lwi == 1 & exp2 == 1, 1, 0)]

# At-risk-of-poverty or social exclusion (AROPE)
data[, arope := ifelse(pov == 1 | dep == 1 | lwi == 1, 1, 0)]

result11 <- vardcros(Y = c("pov", "dep", "arope"),
                    H = "strata", PSU = "PSU", w_final = "rb050",
                    ID_level1 = "db030", ID_level2 = "rb030",
                    Dom = "rb090", Z = NULL, country = "country",
                    period = "year", dataset=data, linratio = FALSE,
                    withperiod = TRUE, netchanges = TRUE,
                    confidence = .95)

data2 <- data[exp2 == 1]
result12 <- vardcros(Y = c("lwi"), H = "strata",
                    PSU = "PSU", w_final = "rb050",
                    ID_level1 = "db030", ID_level2 = "rb030",
                    Dom = "rb090", Z = NULL,
                    country = "country", period = "year",
                    dataset = data2, linratio = FALSE,
                    withperiod = TRUE, netchanges = TRUE,
                    confidence = .95)

### Example 3
data(eusilc)
set.seed(1)
year <- 2011
data <- data.table(rbind(eusilc, eusilc, eusilc, eusilc),
                  rb010 = c(rep(2008, nrow(eusilc)),

```

```

                                rep(2009, nrow(eusilc)),
                                rep(2010, nrow(eusilc)),
                                rep(2011, nrow(eusilc)))
data[, rb020 := "AT"]

data[, u := 1]
data[age < 0, age := 0]
data[, strata := "XXXX"]
PSU <- data[, .N, keyby = "db030"][, N:=NULL]
PSU[, PSU := trunc(runif(nrow(PSU), 0, 100))]
data <- merge(data, PSU, by = "db030", all = TRUE)
thres <- data.table(rb020 = as.character(rep("AT", 4)),
                    thres = c(11406, 11931, 12371, 12791),
                    rb010 = 2008 : 2011)
data <- merge(data, thres, all.x = TRUE, by = c("rb010", "rb020"))
data[is.na(u), u := 0]
data <- data[u == 1]

#####
# T3      #
#####

T3 <- data[rb010 == year - 3]
T3[, strata1 := strata]
T3[, PSU1 := PSU]
T3[, w1 := rb050]
T3[, inc1 := eqIncome]
T3[, rb110_1 := db030]
T3[, pov1 := inc1 <= thres1]
T3 <- T3[, c("rb020", "rb030", "strata", "PSU", "inc1", "pov1"), with = FALSE]

#####
# T2      #
#####

T2 <- data[rb010 == year - 2]
T2[, strata2 := strata]
T2[, PSU2 := PSU]
T2[, w2 := rb050]
T2[, inc2 := eqIncome]
T2[, rb110_2 := db030]
setnames(T2, "thres", "thres2")
T2[, pov2 := inc2 <= thres2]
T2 <- T2[, c("rb020", "rb030", "strata2", "PSU2", "inc2", "pov2"), with = FALSE]

#####
# T1      #
#####

T1 <- data[rb010 == year - 1]
T1[, strata3 := strata]
T1[, PSU3 := PSU]
T1[, w3 := rb050]
T1[, inc3 := eqIncome]
T1[, rb110_3 := db030]

```

```

setnames(T1, "thres", "thres3")
T1[, pov3 := inc3 <= thres3]
T1 <- T1[, c("rb020", "rb030", "strata3", "PSU3", "inc3", "pov3"), with = FALSE]

#####
# T0      #
#####
T0 <- data[rb010 == year]
T0[, PSU4 := PSU]
T0[, strata4 := strata]
T0[, w4 := rb050]
T0[, inc4 := eqIncome]
T0[, rb110_4 := db030]
setnames(T0, "thres", "thres4")
T0[, pov4 := inc4 <= thres4]
T0 <- T0[, c("rb020", "rb030", "strata4", "PSU4", "w4", "inc4", "pov4"), with = FALSE]
apv <- merge(T3, T2, all = TRUE, by = c("rb020", "rb030"))
apv <- merge(apv, T1, all = TRUE, by = c("rb020", "rb030"))
apv <- merge(apv, T0, all = TRUE, by = c("rb020", "rb030"))
apv <- apv[(!is.na(inc1)) & (!is.na(inc2)) & (!is.na(inc3)) & (!is.na(inc4))]
apv[, ppr := ifelse(((pov4 == 1) & ((pov1 == 1 & pov2 == 1 & pov3 == 1) | (pov1 == 1 &
  pov2 == 1 & pov3 == 0) | (pov1 == 1 & pov2 == 0 & pov3 == 1) |
  (pov1 == 0 & pov2 == 1 & pov3 == 1))), 1, 0)]

data[, id2 := .I]
result20 <- vardcros(Y = "ppr", H = "strata", PSU = "PSU",
  w_final = "w4", ID_level1="rb030",
  ID_level2 = "rb030", Dom = NULL,
  Z = NULL, country = "rb020",
  period = NULL, dataset = apv,
  linratio = FALSE,
  withperiod = FALSE,
  netchanges = FALSE,
  confidence = .95)

## End(Not run)

```

vardcrospoor

Variance estimation for cross-sectional, longitudinal measures for indicators on social exclusion and poverty

Description

Computes the variance estimation for cross-sectional and longitudinal measures for indicators on social exclusion and poverty.

Usage

```

vardcrospoor(Y, age = NULL, p1085 = NULL,
  month_at_work = NULL, Y_den = NULL,
  Y_thres = NULL, wght_thres = NULL,

```

```
H, PSU, w_final, ID_level1, ID_level2,
Dom = NULL, country = NULL,
period, sort = NULL, gender = NULL,
dataset = NULL, X = NULL, countryX = NULL,
periodX = NULL, X_ID_level1 = NULL,
ind_gr = NULL, g = NULL, q = NULL,
datasetX = NULL, percentage = 60,
order_quant = 50, alpha = 20,
use.estVar = FALSE, withperiod = TRUE,
netchanges = TRUE, confidence = .95,
outp_lin = FALSE, outp_res = FALSE,
type = "linrmpg", checking = TRUE)
```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
age	Age variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
p1085	Retirement variable (Number of months spent in retirement or early retirement). One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
month_at_work	Variable for total number of month at work (sum of the number of months spent at full-time work as employee, number of months spent at part-time work as employee, number of months spent at full-time work as self-employed (including family worker), number of months spent at part-time work as self-employed (including family worker)). One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Y_den	Denominator variable (for example gross individual earnings). One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Y_thres	Variable (for example equalized disposable income) used for computation and linearization of poverty threshold. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset). Variable specified for <code>inc</code> is used as <code>income_thres</code> if <code>income_thres</code> is not defined.
wght_thres	Weight variable used for computation and linearization of poverty threshold. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number. Variable specified for <code>weight</code> is used as <code>wght_thres</code> if <code>wght_thres</code> is not defined.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
ID_level2	Optional variable for unit ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables are calculated for each domain. An object convertible to data. table or variable names as character vector, column numbers.
country	Variable for the survey countries. The values for each country are computed independently. Object convertible to data. table or variable names as character, column numbers.
period	Variable for the survey periods. The values for each period are computed independently. Object convertible to data. table or variable names as character, column numbers.
sort	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column data. table or variable name as character, column number.
gender	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column data. table or variable name as character, column number.
dataset	Optional survey data object convertible to data. table.
X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to data. table or variable names as character, column numbers.
countryX	Optional variable for the survey countries. The values for each country are computed independently. Object convertible to data. table or variable names as character, column numbers.
periodX	Optional variable of the survey periods and countries. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to data. table or variable names as character, column numbers.
X_ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column data. table or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column data. table or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column data. table or variable name as character, column number.
datasetX	Optional survey data object in household level convertible to data. table.
percentage	A numeric value in range [0, 100] for p in the formula for poverty threshold computation:

$$\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$$

For example, to compute poverty threshold equal to 60% of some income quantile, p should be set equal to 60.

order_quant	A numeric value in range [0, 100] for α in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ <p>For example, to compute poverty threshold equal to some percentage of median income, α should be set equal to 50.</p>
alpha	a numeric value in range [0, 100] for the order of the income quantile share ratio (in percentage).
use.estVar	Logical value. If value is TRUE, then R function <code>estVar</code> is used for the estimation of covariance matrix of the residuals. If value is FALSE, then R function <code>estVar</code> is not used for the estimation of covariance matrix of the residuals.
withperiod	Logical value. If TRUE is value, the results is with period, if FALSE, without period.
netchanges	Logical value. If value is TRUE, then produce two objects: the first object is aggregation of weighted data by period (if available), country, strata and PSU, the second object is an estimation for Y, the variance, gradient for numerator and denominator by country and period (if available). If value is FALSE, then both objects containing NULL.
confidence	Optional positive value for confidence interval. This variable by default is 0.95.
outp_lin	Logical value. If TRUE linearized values of the ratio estimator will be printed out.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.
type	a character vector (of length one unless several.ok is TRUE), example "linarpr", "linarpt", "lingpg", "linpoormed", "linrmpg", "lingini", "lingini2", "lingqr", "linarr", "linr-mir".
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Value

A list with objects are returned by the function:

lin_out	A data.table containing the linearized values of the ratio estimator with ID_level2 and PSU.
res_out	A data.table containing the estimated residuals of calibration with ID_level1 and PSU.
data_net_changes	A data.table containing aggregation of weighted data by period (if available), country, strata, PSU.
results	A data.table containing period - survey periods, country - survey countries, Dom - optional variable of the population domains, type - type variable, count_respondents - the count of respondents, pop_size - the population size (in numbers of individuals), estim - the estimated value,

se - the estimated standard error,
var - the estimated variance,
rse - the estimated relative standard error (coefficient of variation),
cv - the estimated relative standard error (coefficient of variation) in percentage.

References

Guillaume Osier, Yves Berger, Tim Goedeme, (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, Eurostat Methodologies and Working papers, URL <http://ec.europa.eu/eurostat/documents/3888793/5855973/KS-RA-13-024-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

See Also

[linrmir](#), [linarr](#), [vardchanges](#)

Examples

```
data(eusilc)
set.seed(1)

data <- data.table(rbind(eusilc, eusilc),
                  year = c(rep(2010, nrow(eusilc)),
                          rep(2011, nrow(eusilc))))

data[age < 0, age := 0]
PSU <- data[, .N, keyby = "db030"][, N := NULL]
PSU[, PSU := trunc(runif(nrow(PSU), 0, 100))]
PSU$inc <- runif(nrow(PSU), 20, 100000)
data <- merge(data, PSU, all = TRUE, by = "db030")
PSU <- eusilc <- NULL
data[, strata := c("XXXX")]
data$p1085 <- 12 * trunc(runif(nrow(data), 0, 2))
data$month_at_work <- 12 * trunc(runif(nrow(data), 0, 2))
data[, id_l2 := paste0("V", .I)]

result <- vardcrospoor(Y = "inc", age = "age",
                      p1085 = "p1085",
                      month_at_work = "month_at_work",
                      Y_den = "inc", Y_thres = "inc",
                      wght_thres = "rb050",
                      H = "strata", PSU = "PSU",
```

```

w_final="rb050", ID_level1 = "db030",
ID_level2 = "id_l2",
Dom = c("rb090", "db040"),
country = NULL, period = "year",
sort = NULL, gender = NULL,
dataset = data,
percentage = 60,
order_quant = 50L,
alpha = 20,
confidence = 0.95,
type = "linmpg")

## Not run:
result2 <- vardcrospeer(Y = "inc", age = "age",
  pl085 = "pl085",
  month_at_work = "month_at_work",
  Y_den = "inc", Y_thres = "inc",
  wght_thres = "rb050",
  H = "strata", PSU = "PSU",
  w_final = "rb050", ID_level1 = "db030",
  ID_level2 = "id_l2",
  Dom = c("rb090", "db040"),
  period = "year", sort = NULL,
  gender = NULL, dataset = data,
  percentage = 60,
  order_quant = 50L,
  alpha = 20,
  confidence = 0.95,
  type = "linmpg")

## End(Not run)

```

vardom

Variance estimation of the sample surveys in domain by the ultimate cluster method

Description

Computes the variance estimation of the sample surveys in domain by the ultimate cluster method.

Usage

```

vardom(Y, H, PSU, w_final,
  id = NULL, Dom = NULL,
  period = NULL, PSU_sort = NULL,
  N_h = NULL, fh_zero = FALSE,
  PSU_level = TRUE, Z = NULL,
  X = NULL, ind_gr = NULL,
  g = NULL, q = NULL,
  dataset = NULL,
  confidence = .95, percentratio=1,
  outp_lin = FALSE, outp_res = FALSE)

```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, variables of interest are calculated for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, residual estimation of calibration is done independently for each time period. One dimensional object convertible to one-column <code>data.table</code> .
PSU_sort	optional; if <code>PSU_sort</code> is defined, then variance is calculated for systematic sample.
N_h	<p>Number of primary sampling units in population for each stratum (and period if <code>period</code> is not <code>NULL</code>). If <code>N_h = NULL</code> and <code>fh_zero = FALSE</code> (default), <code>N_h</code> is estimated from sample data as sum of weights (<code>w_final</code>) in each stratum (and period if <code>period</code> is not <code>NULL</code>).</p> <p>Optional for single-stage sampling design as it will be estimated from sample data. Recommended for multi-stage sampling design as <code>N_h</code> can not be correctly estimated from the sample data in this case. If <code>N_h</code> is not used in case of multi-stage sampling design (for example, because this information is not available), it is advisable to set <code>fh_zero = TRUE</code>.</p> <p>If <code>period</code> is <code>NULL</code>. A two-column matrix with rows for each stratum. The first column should contain stratum code. The second column - the number of primary sampling units in the population of each stratum.</p> <p>If <code>period</code> is not <code>NULL</code>. A three-column matrix with rows for each intersection of strata and period. The first column should contain period. The second column should contain stratum code. The third column - the number of primary sampling units in the population of each stratum and period.</p>
fh_zero	by default <code>FALSE</code> ; <code>fh</code> is calculated as division of <code>n_h</code> and <code>N_h</code> in each strata, if true, <code>fh</code> value is zero in each strata.
PSU_level	by default <code>TRUE</code> ; if <code>PSU_level</code> is true, in each strata <code>fh</code> is calculated as division of count of PSU in sample (<code>n_h</code>) and count of PSU in frame(<code>N_h</code>). if <code>PSU_level</code> is false, in each strata <code>fh</code> is calculated as division of count of units in sample (<code>n_h</code>) and count of units in frame(<code>N_h</code>), which calculated as sum of weights.
Z	Optional variables of denominator for ratio estimation. Object convertible to <code>data.table</code> or variable names as character, column numbers.

X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
dataset	Optional survey data object convertible to <code>data.table</code> .
confidence	Optional positive value for confidence interval. This variable by default is 0.95.
percentratio	Positive numeric value. All linearized variables are multiplied with <code>percentratio</code> value, by default - 1.
outp_lin	Logical value. If TRUE linearized values of the ratio estimator will be printed out.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.

Details

Calculate variance estimation in domains based on book of Hansen, Hurwitz and Madow.

Value

A list with objects is returned by the function:

lin_out	A <code>data.table</code> containing the linearized values of the ratio estimator with id and PSU.
res_out	A <code>data.table</code> containing the estimated residuals of calibration with id and PSU.
betas	A numeric <code>data.table</code> containing the estimated coefficients of calibration.
all_result	A <code>data.table</code> , which containing variables: <code>variable</code> - names of variables of interest, <code>Dom</code> - optional variable of the population domains, <code>period</code> - optional variable of the survey periods, <code>respondent_count</code> - the count of respondents, <code>pop_size</code> - the estimated size of population, <code>n_nonzero</code> - the count of respondents, who answers are larger than zero, <code>estim</code> - the estimated value, <code>var</code> - the estimated variance, <code>se</code> - the estimated standard error, <code>rse</code> - the estimated relative standard error (coefficient of variation), <code>cv</code> - the estimated relative standard error (coefficient of variation) in percentage, <code>absolute_margin_of_error</code> - the estimated absolute margin of error, <code>relative_margin_of_error</code> - the estimated relative margin of error in percentage,

CI_lower - the estimated confidence interval lower bound,
CI_upper - the estimated confidence interval upper bound,
confidence_level - the positive value for confidence interval,
S2_y_HT - the estimated variance of the y variable in case of total or the estimated variance of the linearised variable in case of the ratio of two totals using non-calibrated weights,
S2_y_ca - the estimated variance of the y variable in case of total or the estimated variance of the linearised variable in case of the ratio of two totals using calibrated weights,
S2_res - the estimated variance of the regression residuals,
var_srs_HT - the estimated variance of the HT estimator under SRS,
var_cur_HT - the estimated variance of the HT estimator under current design,
var_srs_ca - the estimated variance of the calibrated estimator under SRS,
deff_sam - the estimated design effect of sample design,
deff_est - the estimated design effect of estimator,
deff - the overall estimated design effect of sample design and estimator,
n_eff - the effective sample size.

References

Morris H. Hansen, William N. Hurwitz, William G. Madow, (1953), Sample survey methods and theory Volume I Methods and applications, 257-258, Wiley.

Guillaume Osier and Emilio Di Meglio. The linearisation approach implemented by Eurostat for the first wave of EU-SILC: what could be done from the second wave onwards? 2012

Guillaume Osier, Yves Berger, Tim Goedeme, (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, Eurostat Methodologies and Working papers, URL <http://ec.europa.eu/eurostat/documents/3888793/5855973/KS-RA-13-024-EN.PDF>.

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. Survey Methodology, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[domain](#), [lin.ratio](#), [residual_est](#), [vandomh](#), [var_srs](#), [variance_est](#), [variance_othstr](#)

Examples

```

data(eusilc)
dataset <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

aa <- vardomh(Y = "eqIncome", H = "db040", PSU = "db030",
             w_final = "rb050", id = "rb030", Dom = "db040",
             period = NULL, N_h = NULL, Z = NULL,
             X = NULL, g = NULL, q = NULL, dataset = dataset,
             confidence = .95, percentratio = 100,
             outp_lin = TRUE, outp_res = TRUE)

```

vardomh	<i>Variance estimation for sample surveys in domain for one or two stage surveys by the ultimate cluster method</i>
---------	---

Description

Computes the variance estimation in domain for ID_level1.

Usage

```

vardomh(Y, H, PSU, w_final, ID_level1,
        ID_level2 = NULL, Dom = NULL, period = NULL,
        N_h = NULL, PSU_sort = NULL,
        fh_zero = FALSE, PSU_level = TRUE,
        Z = NULL, dataset = NULL,
        X = NULL, periodX = NULL, X_ID_level1 = NULL,
        ind_gr = NULL, g = NULL, q = NULL,
        datasetX = NULL, confidence = .95,
        percentratio = 1, outp_lin = FALSE,
        outp_res = FALSE)

```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ID_level2	Variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

period	Optional variable for the survey periods. If supplied, the values for each period are computed independently. Object convertible to <code>data.table</code> or variable names as character, column numbers.
Dom	Optional variables used to define population domains. If supplied, values are calculated for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
N_h	<p>Number of primary sampling units in population for each stratum (and period if period is not NULL). If <code>N_h = NULL</code> and <code>fh_zero = FALSE</code> (default), <code>N_h</code> is estimated from sample data as sum of weights (<code>w_final</code>) in each stratum (and period if period is not NULL).</p> <p>Optional for single-stage sampling design as it will be estimated from sample data. Recommended for multi-stage sampling design as <code>N_h</code> can not be correctly estimated from the sample data in this case. If <code>N_h</code> is not used in case of multi-stage sampling design (for example, because this information is not available), it is advisable to set <code>fh_zero = TRUE</code>.</p> <p>If period is NULL. A two-column data object convertible to <code>data.table</code> with rows for each stratum. The first column should contain stratum code. The second column - the number of primary sampling units in the population of each stratum.</p> <p>If period is not NULL. A three-column data object convertible to <code>data.table</code> with rows for each intersection of strata and period. The first column should contain period. The second column should contain stratum code. The third column - the number of primary sampling units in the population of each stratum and period.</p>
PSU_sort	optional; if <code>PSU_sort</code> is defined, then variance is calculated for systematic sample.
fh_zero	by default FALSE; <code>fh</code> is calculated as division of <code>n_h</code> and <code>N_h</code> in each strata, if true, <code>fh</code> value is zero in each strata.
PSU_level	by default TRUE; if <code>PSU_level</code> is true, in each strata <code>fh</code> is calculated as division of count of PSU in sample (<code>n_h</code>) and count of PSU in frame (<code>N_h</code>). if <code>PSU_level</code> is false, in each strata <code>fh</code> is calculated as division of count of units in sample (<code>n_h</code>) and count of units in frame (<code>N_h</code>), which calculated as sum of weights.
Z	Optional variables of denominator for ratio estimation. Object convertible to <code>data.table</code> or variable names as character, column numbers or logical vector (length of the vector has to be the same as the column count of dataset).
dataset	Optional survey data object convertible to <code>data.table</code> .
X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
periodX	Optional variable of the survey periods. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
X_ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

g	Optional variable of the g weights. One dimensional object convertible to one-column data. table or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column data. table or variable name as character, column number.
datasetX	Optional survey data object in level1 convertible to data. table.
confidence	Optional positive value for confidence interval. This variable by default is 0.95.
percentratio	Positive numeric value. All linearized variables are multiplied with percentratio value, by default - 1.
outp_lin	Logical value. If TRUE linearized values of the ratio estimator will be printed out.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.

Details

Calculate variance estimation in domains for household surveys based on book of Hansen, Hurwitz and Madow.

Value

A list with objects are returned by the function:

lin_out	A data. table containing the linearized values of the ratio estimator with ID_level2 and PSU.
res_out	A data. table containing the estimated residuals of calibration with ID_level1 and PSU.
betas	A numeric data. table containing the estimated coefficients of calibration.
all_result	A data. table, which containing variables: variable - names of variables of interest, Dom - optional variable of the population domains, period - optional variable of the survey periods, respondent_count - the count of respondents, pop_size - the estimated size of population, n_nonzero - the count of respondents, who answers are larger than zero, estim - the estimated value, var - the estimated variance, se - the estimated standard error, rse - the estimated relative standard error (coefficient of variation), cv - the estimated relative standard error (coefficient of variation) in percentage, absolute_margin_of_error - the estimated absolute margin of error, relative_margin_of_error - the estimated relative margin of error in percentage, CI_lower - the estimated confidence interval lower bound, CI_upper - the estimated confidence interval upper bound, confidence_level - the positive value for confidence interval, S2_y_HT - the estimated variance of the y variable in case of total or the estimated variance of the linearised variable in case of the ratio of two totals using

non-calibrated weights,
S2_y_ca - the estimated variance of the y variable in case of total or the estimated variance of the linearised variable in case of the ratio of two totals using calibrated weights,
S2_res - the estimated variance of the regression residuals,
S2_res - the estimated variance of the regression residuals,
var_srs_HT - the estimated variance of the HT estimator under SRS for household,
var_cur_HT - the estimated variance of the HT estimator under current design for household,
var_srs_ca - the estimated variance of the calibrated estimator under SRS for household,
deff_sam - the estimated design effect of sample design for household,
deff_est - the estimated design effect of estimator for household,
deff - the overall estimated design effect of sample design and estimator for household

References

Morris H. Hansen, William N. Hurwitz, William G. Madow, (1953), Sample survey methods and theory Volume I Methods and applications, 257-258, Wiley.

Guillaume Osier and Emilio Di Meglio. The linearisation approach implemented by Eurostat for the first wave of EU-SILC: what could be done from the second wave onwards? 2012

Guillaume Osier, Yves Berger, Tim Goedeme, (2013), Standard error estimation for the EU-SILC indicators of poverty and social exclusion, Eurostat Methodologies and Working papers, URL <http://ec.europa.eu/eurostat/documents/3888793/5855973/KS-RA-13-024-EN.PDF>.

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. Survey Methodology, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

See Also

[domain](#), [lin.ratio](#), [residual_est](#), [var_srs](#), [variance_est](#)

Examples

```

data(eusilc)
dataset <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)

aa <- vardomh(Y = "eqIncome", H = "db040", PSU = "db030",
             w_final = "rb050", ID_level1 = "db030",
             ID_level2 = "rb030", Dom = "db040", period = NULL,
             N_h = NULL, Z = NULL, dataset = dataset, X = NULL,
             X_ID_level1 = NULL, g = NULL, q = NULL,
             datasetX = NULL, confidence = 0.95, percentratio = 1,
             outp_lin = TRUE, outp_res = TRUE)

## Not run:
dataset2 <- copy(dataset)
dataset$period <- 1
dataset2$period <- 2
dataset <- data.table(rbind(dataset, dataset2))

# by default without using fh_zero (finite population correction)
aa2 <- vardomh(Y = "eqIncome", H = "db040", PSU = "db030",
              w_final = "rb050", ID_level1 = "db030",
              ID_level2 = "rb030", Dom = "db040", period = "period",
              N_h = NULL, Z = NULL, dataset = dataset,
              X = NULL, X_ID_level1 = NULL,
              g = NULL, q = NULL, datasetX = NULL,
              confidence = .95, percentratio = 1,
              outp_lin = TRUE, outp_res = TRUE)

aa2

# without using fh_zero (finite population correction)
aa3 <- vardomh(Y = "eqIncome", H = "db040", PSU = "db030",
              w_final = "rb050", ID_level1 = "db030",
              ID_level2 = "rb030", Dom = "db040",
              period = "period", N_h = NULL, fh_zero=FALSE,
              Z = NULL, dataset = dataset, X = NULL,
              X_ID_level1 = NULL, g = NULL, q = NULL,
              datasetX = NULL, confidence = .95,
              percentratio = 1, outp_lin = TRUE,
              outp_res = TRUE)

aa3

# with using fh_zero (finite population correction)
aa4 <- vardomh(Y="eqIncome", H="db040", PSU="db030", w_final="rb050",
              ID_level1="db030", ID_level2="rb030", Dom = "db040",
              period = "period", N_h = NULL, fh_zero=TRUE,
              Z = NULL, dataset = dataset,
              X = NULL, X_ID_level1 = NULL,
              g = NULL, q = NULL, datasetX = NULL,
              confidence = .95, percentratio = 1,
              outp_lin = TRUE, outp_res = TRUE)

aa4

```

```
## End(Not run)
```

vandom_othstr	<i>Variance estimation for sample surveys in domain by the two stratification</i>
---------------	---

Description

Computes the variance estimation for sample surveys in domain by the two stratification.

Usage

```
vandom_othstr(Y, H, H2, PSU, w_final, id = NULL,
              Dom = NULL, period=NULL, N_h = NULL,
              N_h2, Z = NULL, X = NULL, ind_gr = NULL,
              g = NULL, q = NULL, dataset = NULL,
              confidence = .95, percentratio=1,
              outp_lin=FALSE, outp_res=FALSE)
```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers.
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
H2	The unit new stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
id	Optional variable for unit ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
Dom	Optional variables used to define population domains. If supplied, linearization of the at-risk-of-poverty rate is done for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
period	Optional variable for survey period. If supplied, residual estimation of calibration is done independently for each time period. One dimensional object convertible to one-column <code>data.table</code> .
N_h	optional data object convertible to <code>data.table</code> . If period is supplied, the time period is at the beginning of the object and after time period in the object is stratum. If period is not supplied, the first column in the object is stratum. In the last column is the total of the population in each stratum.

N_h2	optional data object convertible to <code>data.table</code> . If period is supplied, the time period is at the beginning of the object and after time period in the object is new stratum. If period is not supplied, the first column in the object is new stratum. In the last column is the total of the population in each stratum.
Z	optional variables of denominator for ratio estimation. Object convertible to <code>data.table</code> or variable names as character, column numbers.
X	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
dataset	Optional survey data object convertible to <code>data.table</code> .
confidence	Optional positive value for confidence interval. This variable by default is 0.95.
percentratio	Positive numeric value. All linearized variables are multiplied with <code>percentratio</code> value, by default - 1.
outp_lin	Logical value. If TRUE linearized values of the ratio estimator will be printed out.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.

Value

A list with objects are returned by the function:

lin_out	A <code>data.table</code> containing the linearized values of the ratio estimator with id and PSU.
res_out	A <code>data.table</code> containing the estimated residuals of calibration with id and PSU.
betas	A numeric <code>data.table</code> containing the estimated coefficients of calibration.
s2g	A <code>data.table</code> containing the s^2g value.
all_result	A <code>data.table</code> , which containing variables: respondent_count - the count of respondents, pop_size - the estimated size of population, n_nonzero - the count of respondents, who answers are larger than zero, estim - the estimated value, var - the estimated variance, se - the estimated standard error, rse - the estimated relative standard error (coefficient of variation), cv - the estimated relative standard error (coefficient of variation) in percentage, absolute_margin_of_error - the estimated absolute margin of error,

relative_margin_of_error - the estimated relative margin of error in percentage,
 CI_lower - the estimated confidence interval lower bound,
 CI_upper - the estimated confidence interval upper bound,
 confidence_level - the positive value for confidence interval,
 var_srs_HT - the estimated variance of the HT estimator under SRS,
 var_cur_HT - the estimated variance of the HT estimator under current design,
 var_srs_ca - the estimated variance of the calibrated estimator under SRS,
 deff_sam - the estimated design effect of sample design,
 deff_est - the estimated design effect of estimator,
 deff - the overall estimated design effect of sample design and estimator

References

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

M. Liberts. (2004) Non-response Analysis and Bias Estimation in a Survey on Transportation of Goods by Road.

See Also

[domain](#), [lin.ratio](#), [residual_est](#), [vandomh](#), [var_srs](#), [variance_est](#), [variance_othstr](#)

Examples

```
data(eusilc)

# Example 1
eusilc1 <- eusilc[1:1000, ]
dataset <- data.table(IDd = paste0("V", 1:nrow(eusilc1)), eusilc1)
dataset[, db040_2 := get("db040")]
N_h2 <- dataset[, sum(rb050, na.rm = FALSE), keyby = "db040_2"]

aa<-vandom_othstr(Y = "eqIncome", H = "db040", H2 = "db040_2",
  PSU = "db030", w_final = "rb050", id = "rb030",
  Dom = "db040", period = NULL, N_h = NULL,
  N_h2 = N_h2, Z = NULL, X = NULL, g = NULL,
  q = NULL, dataset = dataset, confidence = .95,
  outp_lin = TRUE, outp_res = TRUE)

## Not run:
# Example 2
dataset <- data.table(IDd = 1:nrow(eusilc), eusilc)
dataset[, db040_2 := get("db040")]
N_h2 <- dataset[, sum(rb050, na.rm = FALSE), keyby = "db040_2"]
```

```
aa <- vardom_othstr(Y = "eqIncome", H = "db040", H2 = "db040_2",
  PSU = "db030", w_final = "rb050", id = "rb030",
  Dom = "db040", period = NULL, N_h2 = N_h2,
  Z = NULL, X = NULL, g = NULL, dataset = dataset,
  q = NULL, confidence = .95, outp_lin = TRUE,
  outp_res = TRUE)

## End(Not run)
```

variance_est

Variance estimation for sample surveys by the ultimate cluster method

Description

Computes the variance estimation by the ultimate cluster method.

Usage

```
variance_est(Y, H, PSU, w_final, N_h = NULL,
  fh_zero = FALSE, PSU_level = TRUE,
  PSU_sort = NULL, period = NULL,
  dataset = NULL, msg = "", checking = TRUE)
```

Arguments

- | | |
|---------|--|
| Y | Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers. |
| H | The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number. |
| PSU | Primary sampling unit variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number. |
| w_final | Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number. |
| N_h | Number of primary sampling units in population for each stratum (and period if <code>period</code> is not <code>NULL</code>). If <code>N_h = NULL</code> and <code>fh_zero = FALSE</code> (default), <code>N_h</code> is estimated from sample data as sum of weights (<code>w_final</code>) in each stratum (and period if <code>period</code> is not <code>NULL</code>).

Optional for single-stage sampling design as it will be estimated from sample data. Recommended for multi-stage sampling design as <code>N_h</code> can not be correctly estimated from the sample data in this case. If <code>N_h</code> is not used in case of multi-stage sampling design (for example, because this information is not available), it is advisable to set <code>fh_zero = TRUE</code> .

If <code>period</code> is <code>NULL</code> . A two-column matrix with rows for each stratum. The first column should contain stratum code. The second column - the number of primary sampling units in the population of each stratum. |

	If period is not NULL. A three-column matrix with rows for each intersection of strata and period. The first column should contain period. The second column should contain stratum code. The third column - the number of primary sampling units in the population of each stratum and period.
fh_zero	by default FALSE; fh is calculated as division of n_h and N_h in each strata, if true, fh value is zero in each strata.
PSU_level	by default TRUE; if PSU_level is true, in each strata fh is calculated as division of count of PSU in sample (n_h) and count of PSU in frame (N_h). if PSU_level is false, in each strata fh is calculated as division of count of units in sample (n_h) and count of units in frame (N_h), which calculated as sum of weights.
PSU_sort	optional; if PSU_sort is defined, then variance is calculated for systematic sample.
period	Optional variable for the survey periods. If supplied, the values for each period are computed independently. Object convertible to data.table or variable names as character, column numbers.
dataset	an optional name of the individual dataset data.table.
msg	an optional printed text, when function print error.
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

If we assume that $n_h \geq 2$ for all h , that is, two or more PSUs are selected from each stratum, then the variance of $\hat{\theta}$ can be estimated from the variation among the estimated PSU totals of the variable Z :

$$\hat{V}(\hat{\theta}) = \sum_{h=1}^H (1 - f_h) \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (z_{hi\bullet} - \bar{z}_{h\bullet\bullet})^2,$$

where

- $z_{hi\bullet} = \sum_{j=1}^{m_{hi}} \omega_{hij} z_{hij}$

- $\bar{z}_{h\bullet\bullet} = \frac{\left(\sum_{i=1}^{n_h} z_{hi\bullet} \right)}{n_h}$

- f_h is the sampling fraction of PSUs within stratum
- h is the stratum number, with a total of H strata
- i is the primary sampling unit (PSU) number within stratum h , with a total of n_h PSUs
- j is the household number within cluster i of stratum h , with a total of m_{hi} household
- w_{hij} is the sampling weight for household j in PSU i of stratum h
- z_{hij} denotes the observed value of the analysis variable z for household j in PSU i of stratum h

Value

a data.table containing the values of the variance estimation by totals.

References

Morris H. Hansen, William N. Hurwitz, William G. Madow, (1953), Sample survey methods and theory Volume I Methods and applications, 257-258, Wiley.

Guillaume Osier and Emilio Di Meglio. The linearisation approach implemented by Eurostat for the first wave of EU-SILC: what could be done from the second onwards? 2012

Eurostat Methodologies and Working papers, Standard error estimation for the EU-SILC indicators of poverty and social exclusion, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

See Also

[domain](#), [lin.ratio](#), [linarpr](#), [linarpt](#), [lingini](#), [lingini2](#), [lingpg](#), [linpoomed](#), [linqsr](#), [linrmpg](#), [residual_est](#), [vandom](#), [vandomh](#), [varpoord](#), [variance_othstr](#)

Examples

```
Ys <- rchisq(10, 3)
w <- rep(2, 10)
PSU <- 1 : length(Ys)
H <- rep("Strata_1", 10)

# by default without using fh_zero (finite population correction)
variance_est(Y = Ys, H = H, PSU = PSU, w_final = w)

## Not run:
# without using fh_zero (finite population correction)
variance_est(Y = Ys, H = H, PSU = PSU, w_final = w, fh_zero = FALSE)

# with using fh_zero (finite population correction)
variance_est(Y = Ys, H = H, PSU = PSU, w_final = w, fh_zero = TRUE)

## End(Not run)
```

variance_othstr	<i>Variance estimation for sample surveys by the new stratification</i>
-----------------	---

Description

Computes s^2g and the variance estimation by the new stratification.

Usage

```
variance_othstr(Y, H, H2, w_final,
               N_h = NULL, N_h2,
               period = NULL,
               dataset = NULL,
               checking = TRUE)
```

Arguments

Y	Variables of interest. Object convertible to <code>data.table</code> or variable names as character, column numbers or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset).
H	The unit stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset).
H2	The unit new stratum variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset).
w_final	Weight variable. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset).
N_h	optional; either a <code>data.frame</code> giving the first column - stratum, but the second column - the total of the population in each stratum.
N_h2	optional; either a <code>data.frame</code> giving the first column - new stratum, but the second column - the total of the population in each new stratum.
period	Optional variable for the survey periods. If supplied, the values for each period are computed independently. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number or logical vector with only one TRUE value (length of the vector has to be the same as the column count of dataset).
dataset	Optional survey data object convertible to <code>data.table</code> .
checking	Optional variable if this variable is TRUE, then function checks data preparation errors, otherwise not checked. This variable by default is TRUE.

Details

It is possible to compute population size M_g from sampling frame. The standard deviation of g -th stratum is

$$S_g^2 = \frac{1}{M_g - 1} \sum_{k=1}^{M_g} (y_{gk} - \bar{Y}_g)^2 = \frac{1}{M_g - 1} \sum_{k=1}^{M_g} y_{gk}^2 - \frac{M_g}{M_g - 1} \bar{Y}_g^2$$

$\sum_{k=1}^{M_g} y_{gk}^2$ and \bar{Y}_g^2 have to be estimated to estimate S_g^2 . Estimate of $\sum_{k=1}^{M_g} y_{gk}^2$ is $\sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{gi}^2 z_{hi}$, where $z_{hi} = \begin{cases} 0, & h_i \notin \theta_g \\ 1, & h_i \in \theta_g \end{cases}$, θ_g is the index group of successfully surveyed units belonging to g -th stratum. Estimate of \bar{Y}_g^2 is

$$\hat{\bar{Y}}_g^2 = (\hat{Y}_g)^2 - \text{Var}(\hat{Y}_g)$$

$$\hat{Y}_g = \frac{\hat{Y}_g}{M_g} = \frac{1}{M_g} \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} z_{hi}$$

So the estimate of S_g^2 is

$$s_g^2 = \frac{1}{M_g - 1} \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi}^2 z_{hi} - \frac{M_g}{M_g - 1} \left(\left(\frac{1}{M_g} \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{hi} z_{hi} \right)^2 - \frac{1}{M_g^2} \sum_{h=1}^H N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{hi} z_{hi} - \frac{1}{n_h} \sum_{t=1}^{n_h} y_{ht} z_{ht} \right)^2 \right)$$

Two conditions have to realize to estimate S_g^2 : $n_h > 1, \forall g$ and $\theta_g \neq 0, \forall g$.

Variance of \hat{Y} is

$$\text{Var}(\hat{Y}) = \sum_{g=1}^G M_g^2 \left(\frac{1}{m_g} - \frac{1}{M_g} \right) S_g^2$$

Estimate of $\text{Var}(\hat{Y})$ is

$$\hat{\text{Var}}(\hat{Y}) = \sum_{g=1}^G M_g^2 \left(\frac{1}{m_g} - \frac{1}{M_g} \right) s_g^2$$

Value

A list with objects are returned by the function:

betas	A numeric data . table containing the estimated coefficients of calibration.
s2g	A data . table containing the s^2g value.
var_est	A data . table containing the values of the variance estimation.

References

M. Liberts. (2004) Non-response Analysis and Bias Estimation in a Survey on Transportation of Goods by Road.

See Also

[domain](#), [lin.ratio](#), [linarpr](#), [linarpt](#), [lingini](#), [lingini2](#), [lingpg](#), [linpoomed](#), [linqsr](#), [linrmpg](#), [residual_est](#), [vandom](#), [vandom_othstr](#), [vandomh](#), [varpoord](#)

Examples

```

period=NULL
dataset=NULL
Y <- data.table(matrix(runif(50) * 5, ncol = 5))

H <- data.table(H = as.integer(trunc(5 * runif(10))))
H2 <- data.table(H2 = as.integer(trunc(3 * runif(10))))

N_h <- data.table(matrix(0 : 4, 5, 1))
setnames(N_h, names(N_h), "H")
N_h[, sk:= 10]

N_h2 <- data.table(matrix(0 : 2, 3, 1))
setnames(N_h2, names(N_h2), "H2")
N_h2[, sk2:= 4]

w_final <- rep(2, 10)

vo <- variance_othstr(Y = Y, H = H, H2 = H2,
                     w_final = w_final,
                     N_h = N_h, N_h2 = N_h2,
                     period = NULL,
                     dataset = NULL)

vo
```

varpoord

Estimation of the variance and deff for sample surveys for indicators on social exclusion and poverty

Description

Computes the estimation of the variance for indicators on social exclusion and poverty.

Usage

```

varpoord(Y, w_final, age = NULL, pl085 = NULL, month_at_work = NULL,
        Y_den = NULL, Y_thres = NULL, wght_thres = NULL,
        ID_level1, ID_level2 = NULL, H, PSU, N_h, PSU_sort = NULL,
        fh_zero = FALSE, PSU_level = TRUE, sort = NULL, Dom = NULL,
```

```

period = NULL, gender = NULL, dataset = NULL, X = NULL, periodX = NULL,
X_ID_level1 = NULL, ind_gr = NULL, g = NULL, q = NULL, datasetX = NULL,
percentage = 60, order_quant = 50, alpha = 20, confidence = 0.95,
  outp_lin = FALSE, outp_res = FALSE, kern_method = "gaussian",
  r = NULL, ro = NULL, h_breaks = NULL, type = "linrmpg")

```

Arguments

Y	Study variable (for example equalized disposable income or gross pension income). One dimensional object convertible to one-column data. table or variable name as character, column number.
w_final	Weight variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
age	Age variable. One dimensional object convertible to one-column data. frame or variable name as character, column number.
pl085	Retirement variable (Number of months spent in retirement or early retirement). One dimensional object convertible to one-column data. table or variable name as character, column number.
month_at_work	Variable for total number of month at work (sum of the number of months spent at full-time work as employee, number of months spent at part-time work as employee, number of months spent at full-time work as self-employed (including family worker), number of months spent at part-time work as self-employed (including family worker)). One dimensional object convertible to one-column data. table or variable name as character, column number.
Y_den	Denominator variable (for example gross individual earnings). One dimensional object convertible to one-column data. table or variable name as character, column number.
Y_thres	Variable (for example equalized disposable income) used for computation and linearization of poverty threshold. One dimensional object convertible to one-column data. table or variable name as character, column number. Variable specified for inc is used as income_thres if income_thres is not defined.
wght_thres	Weight variable used for computation and linearization of poverty threshold. One dimensional object convertible to one-column data. table or variable name as character, column number. Variable specified for weight is used as wght_thres if wght_thres is not defined.
ID_level1	Variable for level1 ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
ID_level2	Optional variable for unit ID codes. One dimensional object convertible to one-column data. table or variable name as character, column number.
H	The unit stratum variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
PSU	Primary sampling unit variable. One dimensional object convertible to one-column data. table or variable name as character, column number.
N_h	Number of primary sampling units in population for each stratum (and period if period is not NULL). If N_h = NULL and fh_zero = FALSE (default), N_h is

estimated from sample data as sum of weights (`w_final`) in each stratum (and period if period is not NULL).

Optional for single-stage sampling design as it will be estimated from sample data. Recommended for multi-stage sampling design as `N_h` can not be correctly estimated from the sample data in this case. If `N_h` is not used in case of multi-stage sampling design (for example, because this information is not available), it is advisable to set `fh_zero = TRUE`.

If period **is** NULL. A two-column data object convertible to `data.table` with rows for each stratum. The first column should contain stratum code. The second column - the number of primary sampling units in the population of each stratum.

If period **is not** NULL. A three-column data object convertible to `data.table` with rows for each intersection of strata and period. The first column should contain period. The second column should contain stratum code. The third column - the number of primary sampling units in the population of each stratum and period.

<code>PSU_sort</code>	optional; if <code>PSU_sort</code> is defined, then variance is calculated for systematic sample.
<code>fh_zero</code>	by default FALSE; <code>fh</code> is calculated as division of <code>n_h</code> and <code>N_h</code> in each strata, if true, <code>fh</code> value is zero in each strata.
<code>PSU_level</code>	by default TRUE; if <code>PSU_level</code> is true, in each strata <code>fh</code> is calculated as division of count of PSU in sample (<code>n_h</code>) and count of PSU in frame(<code>N_h</code>). if <code>PSU_level</code> is false, in each strata <code>fh</code> is calculated as division of count of units in sample (<code>n_h</code>) and count of units in frame(<code>N_h</code>), which calculated as sum of weights.
<code>sort</code>	Optional variable to be used as tie-breaker for sorting. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
<code>Dom</code>	Optional variables used to define population domains. If supplied, variables is calculated for each domain. An object convertible to <code>data.table</code> or variable names as character vector, column numbers.
<code>period</code>	Optional variable for survey period. If supplied, variables is calculated for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
<code>gender</code>	Numerical variable for gender, where 1 is for males, but 2 is for females. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.
<code>dataset</code>	Optional survey data object convertible to <code>data.frame</code> .
<code>X</code>	Optional matrix of the auxiliary variables for the calibration estimator. Object convertible to <code>data.table</code> or variable names as character, column numbers.
<code>periodX</code>	Optional variable of the survey periods. If supplied, residual estimation of calibration is done independently for each time period. Object convertible to <code>data.table</code> or variable names as character, column numbers.
<code>X_ID_level1</code>	Variable for level1 ID codes. One dimensional object convertible to one-column <code>data.table</code> or variable name as character, column number.

ind_gr	Optional variable by which divided independently X matrix of the auxiliary variables for the calibration. One dimensional object convertible to one-column data.table or variable name as character, column number.
g	Optional variable of the g weights. One dimensional object convertible to one-column data.table or variable name as character, column number.
q	Variable of the positive values accounting for heteroscedasticity. One dimensional object convertible to one-column data.table or variable name as character, column number.
datasetX	Optional survey data object in household level convertible to data.table.
percentage	A numeric value in range [0, 100] for p in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to 60% of some income quantile, p should be set equal to 60.
order_quant	A numeric value in range [0, 100] for α in the formula for poverty threshold computation: $\frac{p}{100} \cdot Z_{\frac{\alpha}{100}}.$ For example, to compute poverty threshold equal to some percentage of median income, α should be set equal to 50.
alpha	a numeric value in range [0, 100] for the order of the income quantile share ratio (in percentage).
confidence	Optional positive value for confidence interval. This variable by default is 0.95.
outp_lin	Logical value. If TRUE linearized values of the ratio estimator will be printed out.
outp_res	Logical value. If TRUE estimated residuals of calibration will be printed out.
kern_method	A character specifying the method used for density calculation.
r	Value specifying the smoothing spline degree.
ro	Value specifying the smoothing spline error from histogram.
h_breaks	Value specifying the histogram bins.
type	a character vector (of length one unless several.ok is TRUE), example "linarpr", "linarpt", "lingpg", "linpoormed", "linrmpg", "lingini", "lingini2", "linqsr", "linarr", "linr-mir".

Value

A list with objects are returned by the function:

lin_out	A data.table containing the linearized values of the ratio estimator with ID_level2 and PSU.
res_out	A data.table containing the estimated residuals of calibration with ID_level1 and PSU.
betas	A numeric data.table containing the estimated coefficients of calibration.

all_result A data.table, which containing variables:

- respondent_count - the count of respondents,
- pop_size - the estimated size of population,
- n_nonzero - the count of respondents, who answers are larger than zero,
- value - the estimated value,
- var - the estimated variance,
- se - the estimated standard error,
- rse - the estimated relative standard error (coefficient of variation),
- cv - the estimated relative standard error (coefficient of variation) in percentage,
- absolute_margin_of_error - the estimated absolute margin of error,
- relative_margin_of_error - the estimated relative margin of error in percentage,
- CI_lower - the estimated confidence interval lower bound,
- CI_upper - the estimated confidence interval upper bound,
- confidence_level - the positive value for confidence interval,
- S2_y_HT - the estimated variance of the y variable in case of total or the estimated variance of the linearised variable in case of the ratio of two totals using non-calibrated weights,
- S2_y_ca - the estimated variance of the y variable in case of total or the estimated variance of the linearised variable in case of the ratio of two totals using calibrated weights,
- S2_res - the estimated variance of the regression residuals,
- var_srs_HT - the estimated variance of the HT estimator under SRS for household,
- var_cur_HT - the estimated variance of the HT estimator under current design for household,
- var_srs_ca - the estimated variance of the calibrated estimator under SRS for household,
- deff_sam - the estimated design effect of sample design for household,
- deff_est - the estimated design effect of estimator for household,
- deff - the overall estimated design effect of sample design and estimator for household

References

Eric Graf and Yves Tille, Variance Estimation Using Linearization for Poverty and Social Exclusion Indicators, *Survey Methodology*, June 2014 61 Vol. 40, No. 1, pp. 61-79, Statistics Canada, Catalogue no. 12-001-X, URL <http://www.statcan.gc.ca/pub/12-001-x/12-001-x2014001-eng.pdf>

Guillaume Osier and Emilio Di Meglio. The linearisation approach implemented by Eurostat for the first wave of EU-SILC: what could be done from the second wave onwards? 2012

Guillaume Osier (2009). Variance estimation for complex indicators of poverty and inequality. *Journal of the European Survey Research Association*, Vol.3, No.3, pp. 167-195, ISSN 1864-3361, URL <http://ojs.ub.uni-konstanz.de/srm/article/view/369>.

Eurostat Methodologies and Working papers, Standard error estimation for the EU-SILC indicators of poverty and social exclusion, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

Jean-Claude Deville (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, 25, 193-203, URL <http://www.statcan.gc.ca/pub/12-001-x/1999002/article/4882-eng.pdf>.

Eurostat Methodologies and Working papers, Handbook on precision requirements and variance estimation for ESS household surveys, 2013, URL <http://ec.europa.eu/eurostat/documents/3859598/5927001/KS-RA-13-029-EN.PDF>.

MATTI LANGEL - YVES TILLE, Corrado Gini, a pioneer in balanced sampling and inequality theory. *METRON - International Journal of Statistics*, 2011, vol. LXIX, n. 1, pp. 45-65, URL <ftp://metron.sta.uniroma1.it/RePEc/articoli/2011-1-3.pdf>.

Morris H. Hansen, William N. Hurwitz, William G. Madow, (1953), *Sample survey methods and theory Volume I Methods and applications*, 257-258, Wiley.

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

Working group on Statistics on Income and Living Conditions (2004) Common cross-sectional EU indicators based on EU-SILC; the gender pay gap. *EU-SILC 131-rev/04*, Eurostat.

See Also

[varodom](#), [varodomh](#), [linarpt](#)

Examples

```
data(eusilc)
dataset <- data.table(IDd = paste0("V", 1 : nrow(eusilc)), eusilc)
dataset1 <- dataset[1 : 1000]

# use dataset1 by default without using fh_zero (finite population correction)
aa<-varpoord(Y = "eqIncome", w_final = "rb050",
            Y_thres = NULL, wght_thres = NULL,
            ID_level1 = "db030", ID_level2 = "IDd",
            H = "db040", PSU = "rb030", N_h = NULL,
            sort = NULL, Dom = NULL,
            gender = NULL, X = NULL,
            X_ID_level1 = NULL, g = NULL,
            q=NULL, datasetX = NULL,
            dataset = dataset1, percentage = 60,
```

```

        order_quant = 50L, alpha = 20,
        confidence = .95, outp_lin = FALSE,
        outp_res = FALSE, type = "linarpt")
aa

## Not run:
# use dataset1 by default with using fh_zero (finite population correction)
aa2 <- varpoord(Y = "eqIncome", w_final = "rb050",
  Y_thres = NULL, wght_thres = NULL,
  ID_level1 = "db030", ID_level2 = "IDd",
  H = "db040", PSU = "rb030", N_h = NULL,
  fh_zero = TRUE, sort = NULL, Dom = "db040",
  gender = NULL, X = NULL, X_ID_level1 = NULL,
  g = NULL, datasetX = NULL, dataset = dataset1,
  percentage = 60, order_quant = 50L,
  alpha = 20, confidence = .95, outp_lin = FALSE,
  outp_res = FALSE, type = "linarpt")

aa2
aa2$all_result

# using dataset
aa4 <- varpoord(Y = "eqIncome", w_final = "rb050",
  Y_thres = NULL, wght_thres = NULL,
  ID_level1 = "db030", ID_level2 = "IDd",
  H = "db040", PSU = "rb030", N_h = NULL,
  sort = NULL, Dom = "db040",
  gender = NULL, X = NULL,
  X_ID_level1 = NULL, g = NULL,
  datasetX = NULL, dataset = dataset,
  percentage = 60, order_quant = 50L,
  alpha = 20, confidence = .95,
  outp_lin = TRUE, outp_res = TRUE,
  type = "linarpt")
aa4$lin_out[20 : 40]

## End(Not run)

```

var_srs

The estimation of the simple random sampling.

Description

Computes the estimation of the simple random sampling.

Usage

```
var_srs(Y, w)
```

Arguments

Y	The variables of interest.
w	Weight variable. One dimensional object convertible to one-column data.frame.

Value

A list with objects are returned by the function:

S2p	A data.table containing the values of the variance estimation of the population.
varsrs	A data.table containing the values of the variance estimation of the simple random sampling.

References

Yves G. Berger, Tim Goedeme, Guillame Osier (2013). Handbook on standard error estimation and other related sampling issues in EU-SILC, URL https://ec.europa.eu/eurostat/cros/content/handbook-standard-error-estimation-and-other-related-sampling-issues-ver-29072013_en

See Also

[vandom](#), [vandomh](#), [varpoord](#)

Examples

```
Ys <- matrix(rchisq(10, 3), 10, 1)
ws <- c(rep(2, 5), rep(3, 5))
var_srs(Ys, ws)
```

Index

- *Topic **Linearization**
 - incPercentile, 6
 - linarpr, 8
 - linarpt, 10
 - linarr, 13
 - lingini, 15
 - lingini2, 17
 - lingpg, 19
 - linpoormed, 21
 - linqsr, 23
 - linrmir, 25
 - linrmpg, 27
 - *Topic **package**
 - vardpoor-package, 2
 - *Topic **survey**
 - domain, 4
 - lin.ratio, 7
 - residual_est, 29
 - *Topic **vardchangannual**
 - vardannual, 31
 - *Topic **vardchanges**
 - vardchanges, 38
 - vardchangespoor, 43
 - *Topic **vardchangstrs**
 - vardchangstrs, 48
 - *Topic **vardcros**
 - vardcros, 50
 - vardcrospoor, 57
 - *Topic **vardpoor**
 - vardom, 62
 - vardom_othstr, 71
 - vardomh, 66
 - variance_est, 74
 - variance_othstr, 77
 - *Topic **variance**
 - var_srs, 85
 - *Topic **varpoord**
 - varpoord, 79
- domain, 4, 7, 30, 36, 41, 47, 50, 54, 65, 69, 73,
- 76, 79
- incPercentile, 6, 12, 25
- lin.ratio, 7, 30, 54, 65, 69, 73, 76, 79
- linarpr, 7, 8, 12, 29, 30, 76, 79
- linarpt, 7, 10, 10, 23, 29, 30, 76, 79, 84
- linarr, 13, 61
- lingini, 15, 18, 20, 30, 76, 79
- lingini2, 17, 17, 30, 76, 79
- lingpg, 19, 30, 76, 79
- linpoormed, 21, 29, 30, 76, 79
- linqsr, 7, 17, 18, 20, 23, 30, 76, 79
- linrmir, 25, 61
- linrmpg, 23, 27, 30, 76, 79
- residual_est, 29, 65, 69, 73, 76, 79
- var_srs, 65, 69, 73, 85
- vardannual, 7, 31
- vardchanges, 7, 36, 38, 47, 61
- vardchangespoor, 10, 12, 15, 17, 18, 20, 23, 25, 27, 29, 41, 43
- vardchangstrs, 48
- vardcros, 7, 36, 41, 47, 50
- vardcrospoor, 10, 12, 15, 17, 18, 20, 23, 25, 27, 29, 47, 57
- vardom, 5, 7, 30, 62, 76, 79, 84, 86
- vardom_othstr, 71, 79
- vardomh, 5, 7, 30, 65, 66, 73, 76, 79, 84, 86
- vardpoor (vardpoor-package), 2
- vardpoor-package, 2
- variance_est, 30, 65, 69, 73, 74
- variance_othstr, 30, 65, 73, 76, 77
- varpoord, 10, 12, 15, 17, 18, 20, 23, 25, 27, 29, 30, 76, 79, 79, 86