Poverty measures

.....possible target parameters

Head Count Ratio P_0

- measures the proportion of the population that is poor. It is popular because it is easy to understand and measure.

Poverty Gap

- measures the extent to which individuals fall below the poverty line (the poverty gaps) as a proportion of the poverty line

P_0- HCR - ARPR

The *headcount index* (P_0) measures the proportion of the population that is poor.

$$P_0 = \frac{N_p}{N},\qquad P_0 = \frac{1}{N}\sum_{i=1}^{N} I(y_i < z).$$

1.1 At-risk-of-poverty rate (after social transfers)

1.1.1 Definition

The share of persons with an equivalised total net income below 60% national median income.

P_0

1. the headcount index does not take the intensity of poverty into account.

Headcount Poverty Rates in A and B, Assuming Poverty Line of 125

	Expenditure for each individual in country				Headcount poverty rate (P_0)
Expenditure in country A	100	100	150	150	50%
Expenditure in country B	124	124	150	150	50%

2. the headcount index does not indicate how poor the poor are, and hence does not change if people below the poverty line become poorer.

P_0

3. the poverty estimates should be calculated for individuals, not households.

- If 20 percent of households are poor, it may be that 25 percent of the population is poor (if poor households are large) or 15 percent is poor (if poor households are small)

P_0

- survey data are almost always related to households, so to measure poverty at the individual level we must make a critical assumption that all members of a given household enjoy the same level of well-being
- In reality, consumption is not always evenly shared across household members..

P_1

define the poverty gap (G_i) as the poverty line (z) less actual income (y_i) for poor individuals; the gap is considered to be zero for everyone else.

$$G_i = (z - y_i) \times I(y_i < z)$$

- The sum of these poverty gaps gives the minimum cost of eliminating poverty, if transfers were perfectly targeted.

$$P_1 = \frac{1}{N} \sum_{i=1}^{N} \frac{G_i}{z}.$$

P_1

Calculating the Poverty Gap Index, Assuming Poverty Line of 125

	Expenditure for each individual in country				Poverty gap index (P_1)
Expenditure in country C	100	110	150	160	
Poverty gap	25	15	0	0	
G_i/z	0.20	0.12	0	0	0.08 [= 0.32/4]

The minimum cost of eliminating poverty using targeted transfers is simply the sum of all the poverty gaps in a population; every gap is filled up to the poverty line. However, this interpretation is only reasonable if the transfers could be made perfectly efficiently, for instance, with lump sum transfers, which is implausible.

Estimation for planned domains - 1

Sample is divided into subsamples s_d, $d = 1,\ldots,D$

Planned domains:

Stratified sampling with domains = strata

- The population domains U_d can be regarded as separate subpopulations
- Domain sizes N_d in domains U_d are assumed known
- Sample size n_d in domain sample $s_d \subset U_d$ is fixed in advance
- **Standard population estimators are applicable as such** as the Horvitz-thompson estimator

Estimation for planned domains - 3

NOTES

Stratified sampling with a suitable *allocation scheme* (e.g. optimal (Neyman) or power (Bankier) allocation) is advisable in practical applications, in order to obtain control over domain sample sizes

Singh, Gambino and Mantel (1994) describe allocation strategies to attain reasonable accuracy for small domains, still retaining good accuracy for large domains

Horvitz-Thompson estimator of domain totals

Horvitz-Thompson (HT) estimator (*expansion estimator*) is the basic *design-based direct* estimator of the domain total $t_d = \sum_{k \in U_d} y_k$, $d = 1,\ldots,D$:

$$\hat{t}_{dHT} = \sum_{k \in U_d} I_k y_k / \pi_k = \sum_{k \in s_d} y_k / \pi_k = \sum_{k \in s_d} a_k y_k \qquad (1)$$

HT estimates of domain totals are additive: they sum up to the HT estimator $\hat{t}_{HT} = \sum_{k \in s} a_k y_k$ of the population total $t = \sum_{k \in U} y_k$

As $E(I_k) = \pi_k$, the HT estimator is design unbiased for t_d

Variance estimation for HT - 1

.....possible target parameters

Standard *variance estimator* for $\hat{t}_{dHT}$ under **planned** domains:

$$\hat{V}\left(\hat{t}_{dHT}\right) = \sum_{k \in s_d} \sum_{l \in s_d} (a_k a_l - a_{kl}) y_k y_l \qquad (2)$$

An alternative Sen-Yates-Grundy formula:

$$\hat{V}\left(\hat{t}_{dHT}\right) = \sum_{k \in s_d} \sum_{l<k; l \in s_d} \left(\frac{a_{kl}}{a_k a_l} - 1\right)(a_k y_k - a_l y_l)^2 \qquad (3)$$

NOTE: Both (2) and (3) are somewhat impractical... Why?

Variance estimation for HT - 3

Variance estimation for planned domains in practice

$$\hat{V}_A\left(\hat{t}_{dHT}\right) = \frac{1}{n_d(n_d - 1)} \sum_{k \in s_d} \left(n_d a_k y_k - \hat{t}_{dHT}\right)^2 \qquad (4)$$

For example, SAS Procedure SURVEYMEANS uses (4)

Variance estimation for HT - 3

Unplanned domains:

Variance estimator should account for random domain sizes

Approximate variance estimator by using *extended domain variables* y_{dk}:

$$\hat{V}_U\left(\hat{t}_{dHT}\right) = \frac{1}{n(n-1)}\sum_{k\in s}\left(na_k y_{dk} - \hat{t}_{dHT}\right)^2, \qquad (5)$$

where n is the total sample size

NOTE: e.g. SAS procedure SURVEYMEANS uses (5)

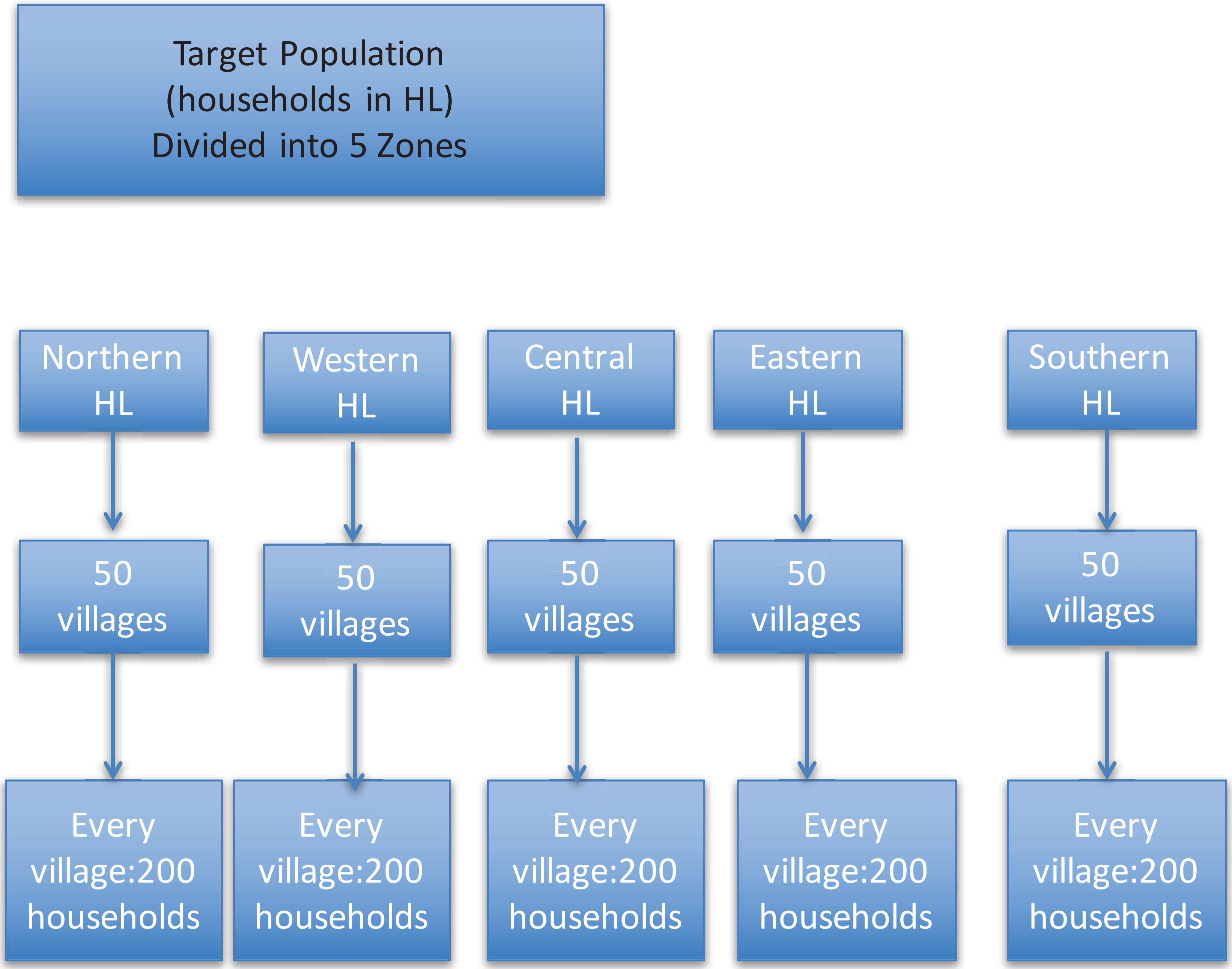
NOTE: Extended domain variables are $y_{dk} = I\{k \in U_d\} y_k$

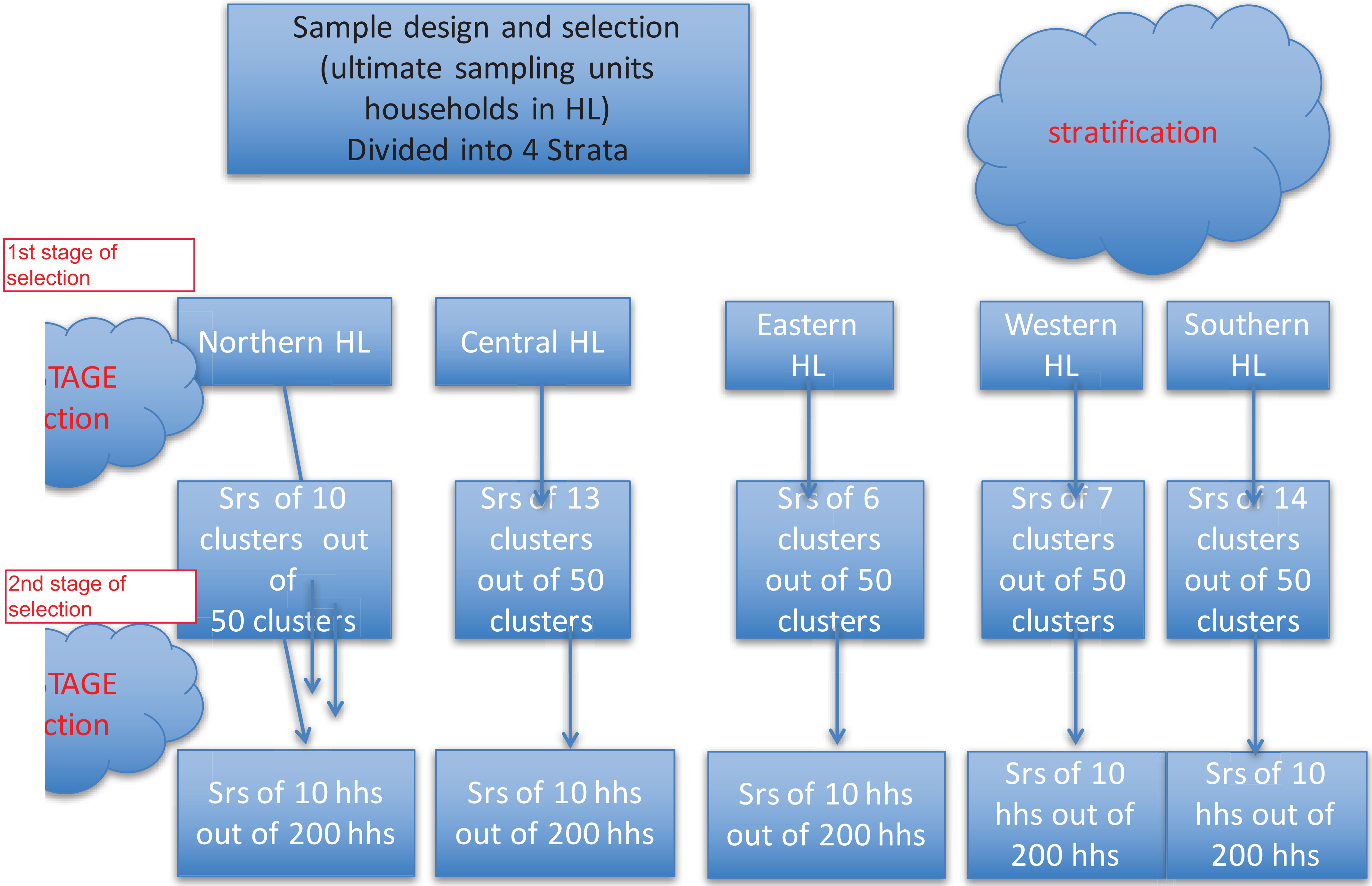
Recall: $y_{dk} = y_k$ if $k \in U_d$, 0 otherwise

Example: estimation for domains

Happy Land Food Survey
Stratified two stage sample survey
H=5 strata; N=50000 households
A= 200 villages in HL (clusters)
a=50 sampled villages

- n=700 households
- Complete coverage of the target population
- Full response of the interviewed households


Target Population
(households in HL)
Divided into 5 Zones
Northern HL
Western HL
Central HL
Eastern HL
Southern HL
50 villages
50 villages
50 villages
50 villages
50 villages
Every village:200 households
Every village:200 households
Every village:200 households
Every village:200 households
Every village:200 households


Sample design and selection
(ultimate sampling units
households in HL)
Divided into 4 Strata
stratification
1st stage of selection
2nd stage of selection
Northern HL
Central HL
Eastern HL
Western HL
Southern HL
Srs of 10 clusters out of 50 clusters
Srs of 13 clusters out of 50 clusters
Srs of 6 clusters out of 50 clusters
Srs of 7 clusters out of 50 clusters
Srs of 14 clusters out of 50 clusters
Srs of 10 hhs out of 200 hhs
Srs of 10 hhs out of 200 hhs
Srs of 10 hhs out of 200 hhs
Srs of 10 hhs out of 200 hhs
Srs of 10 hhs out of 200 hhs

***Exercise*:**
fill it using the probabilities of inclusions!

H code	Sa-weights	village	stratum	FoodExp	H-size	H-income	HH-educ	HH-age
1		1	north HL	12	1	2500	1	56
2		1	north HL	16,5	1	2800	1	70
3		1	north HL	18	1	2000	1	20
4		1	north HL	17	1	4500	1	60
5		1	north HL	46,5	1	8000	1	40
6		1	north HL	45	1	7000	1	51
7		1	north HL	15	1	3500	1	76
8		1	north HL	60	2	2800	1	20
9		1	north HL	15	2	2500	1	51
10		1	north HL	18	2	4000	1	32
11		2	north HL	22,5	2	5000	1	47
12		2	north HL	20	2	8000	1	35
13		2	north HL	97	2	5500	1	58
14		2	north HL	57	2	6000	1	27
15		2	north HL	39	2	3000	1	38
16		2	north HL	30	2	4000	1	40
17		2	north HL	42	2	3000	1	19

Probability of inclusion of *k-th* hh

This is what I need: $1/\pi_k = a_k$

Sampling weight for the *k-th* household

$\pi_k = \pi_{hi} \times \pi_{hk|i}$

π_{hi}

probability of inclusion of the village (*h-th* stratum)

$\pi_{hk|i}$

probability of inclusion of the household, given that the *i-th* village is included (*h-th* stratum)

Probability of inclusion of *k-th* hh

This is what I need: $1/\pi_k = a_k$

Sampling weight for the *k-th* household

$\pi_k = \pi_{hi} \times \pi_{hk|i}$

π_{hi}

probability of inclusion of the village (*h-th* stratum)

$\pi_{hk|i}$

probability of inclusion of the household, given that the *i-th* village is included (*h-th* stratum)

Package "laeken"

- # Package "laeken"
-
- # Estimation of Laeken indicators using synthetic EU-SILC data
-
- # description of the EU-SILC survey:
- # http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions
-
- # load the library (necessary to install it in R the first time it is used)
- library(laeken)
-
- # load synthetic Austrian EU-SILC data
- data(eusilc)
- dim(eusilc)

Package "laeken"

\# A data frame with 14827 observations on the following 28 variables:

\# db030 integer; the household ID.

\# hsize integer; the number of persons in the household.

\# db040 factor; the federal state in which the household is located (levels Burgenland, Carinthia, Lower Austria, Salzburg, Styria, Tyrol, Upper Austria, Vienna and Vorarlberg).

\# rb030 integer; the personal ID.

- .
- .
- .

\# db090 numeric; the household sample weights.

\# rb050 numeric; the personal sample weights.

Package "laeken"

- # AT-RISK-OF-POVERTY RATE
-
- # at-risk-of-poverty rate: national level
- arpr(eusilc$eqIncome, weights = eusilc$rb050, design =eusicl$db040)
-
- # at-risk-of-poverty rate: federal states level (NUTS 2)
- arpr(eusilc$eqIncome, weights = eusilc$rb050, design =eusilc$db040, breakdown = eusilc$db040)
  - Federal state indicator

Package "laeken"

computing confidence intervals: national level

a <- arpr(eusilc$eqIncome, weights = eusilc$rb050, design =eusicl$db040)

bootVar(inc=eusilc$eqIncome, weights = eusilc$rb050, design =eusilc$db040, indicator=a, R=1000, bootType="naive", ciType="perc")

•

computing confidence intervals: federal states level (NUTS 2)

a.states <- arpr(eusilc$eqIncome, weights = eusilc$rb050, design=eusilc$db040, breakdown=eusilc$db040)

bootVar(inc=eusilc$eqIncome, weights = eusilc$rb050, design =eusilc$db040, indicator=a.states, breakdown=eusilc$db040, R=1000, bootType="naive", ciType="perc")