SUPPLEMENT (to Topic 3)
SAE course 27.1.2015 Risto Lehtonen

Empirical examination of theoretical properties of estimators by simulation experiments

In simulation experiments we draw a large number *J* of independent *n* element samples from *N* element fixed and finite population *U* under the given sampling design $p(s)$.

Parameter $t = \sum_{k \in U} y_k$

Estimator (e.g. HT) $\hat{t} = \sum_{k \in s} y_k / ®_k = \sum_{k \in s} a_k y_k$

Design bias

$$Bias(\hat{t}) = E(\hat{t}) - t \approx \sum_{j=1}^{J} \hat{t}_j / J - t$$

where $\hat{t}_j$ is estimated total from simulated sample *j*

Absolute relative bias $\mathrm{ARB}(\hat{t}) = \left| \sum_{j=1}^{J} \hat{t}_j / J - t \right| / t$

Precision

$$Var(\hat{t}) = E(\hat{t} - E(\hat{t}))^2$$

$$\approx \sum_{j=1}^{J} (\hat{t}_j - \sum_{j=1}^{J} \hat{t}_j / J)^2 / (J - 1)$$

Accuracy

$$MSE(\hat{t}) = E(\hat{t} - t)^2 = Var(\hat{t}) + Bias^2(\hat{t})$$

$$\approx \sum_{j=1}^{J} (\hat{t}_j - t)^2 / J$$

In domain estimation / SAE we can do this for the domains.

Numerical example

Population N = 966 elements
D = 10 domains
Sample size: 25% sample from each domain
Planned domains case
Domains = the strata

			Simulation results, J=500 SRSWOR samples				
domain	**n_pop**	**t_pop**	**Total_estimate**	**Monte Carlo Variance**	**Monte Carlo Std_Error**	**Mean_bias_HT**	**ARB_HT**
1	69	1299.27	1295.5	1508.63	38.84	-3.73022	-0.29
2	120	2532.79	2538.2	3346.02	57.84	5.45510	0.22
3	94	1839.14	1838.4	2695.71	51.92	-0.69534	-0.04
4	86	1864.56	1864.8	1506.47	38.81	0.21536	0.01
5	86	1737.94	1738.6	2596.17	50.95	0.66184	0.04
6	204	4662.57	4660.3	5920.51	76.94	-2.27779	-0.05
7	46	835.20	836.23	1367.70	36.98	1.02921	0.12
8	47	1022.06	1021.4	742.82	27.25	-0.67869	-0.07
9	40	884.18	882.90	595.13	24.40	-1.28033	-0.14
10	174	3593.91	3596.0	3652.72	60.44	2.04726	0.06

Distribution of estimates, d = 10