Helsingin yliopisto
Matematiikan ja tilastotieteen laitos
Risto Lehtonen

SAE-kurssi
2. demot 19.3.2009

HT-estimaattorin varianssin estimointi

(1) Suunniteltu domainrakenne
(*Planned domains*)

Osajoukkojen otoskoot n_d on kiinnitetty ositetun otanta-asetelman mukaisesti.

Oletetaan ositettu SRSWOR = STR-SRSWOR

Otos s_d kokoa n_d alkiota poimitaan **ositteesta** U_d, jossa on N_d alkiota

Asetelmapainot ovat $w_k = N_d / n_d$ kaikille $k \in U_d$

Domain-totaalin $T_d = \sum_{k \in U_d} Y_k$ **HT-estimaattori** on

$$\hat{t}_{dHT} = \sum_{k \in s_d} w_k y_k$$

HT-estimaattorin varianssiestimaattori:

$$\hat{v}_{str-srs}(\hat{t}_{dHT}) = N_d^2 (1 - \frac{n_d}{N_d})(\frac{1}{n_d})\hat{s}_d^2$$

missä $\hat{s}_d^2 = \sum_{k \in s_d} \frac{(y_k - \bar{y}_d)^2}{n_d - 1}$ on tulosmuuttujan varianssi

domainissa d

Total

If you specify the keyword SUM, the procedure computes the estimate of the population total from the survey data. The estimate of the total is the weighted sum over the sample.

$$\widehat{Y} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij} \, y_{hij}$$

For a categorical variable level, $\widehat{Y}$ estimates its total frequency in the population.

Variance and Standard Deviation of the Total

When you specify the keyword STD or the keyword SUM, the procedure estimates the standard deviation of the total. The keyword VARSUM requests the variance of the total.

PROC SURVEYMEANS estimates the variance of the total as

$$\widehat{V}(\widehat{Y}) = \sum_{h=1}^{H} \widehat{V_h}(\widehat{Y})$$

where if n_h>1,

$$\begin{aligned}
\widehat{V_h}(\widehat{Y}) &= \frac{n_h(1-f_h)}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi\cdot} - \bar{y}_{h\cdot\cdot})^2 \\
y_{hi\cdot} &= \sum_{j=1}^{m_{hi}} w_{hij} \, y_{hij} \\
\bar{y}_{h\cdot\cdot} &= \left(\sum_{i=1}^{n_h} y_{hi\cdot} \right) / \, n_h
\end{aligned}$$

and if n_h=1,

$$\widehat{V_h}(\widehat{Y}) = \begin{cases} \text{missing} & \text{if } n_{h'} = 1 \text{ for } h' = 1, 2, \ldots, H \\ 0 & \text{if } n_{h'} > 1 \text{ for some } 1 < h' < H \end{cases}$$

The standard deviation of the total equals