Selection Probabilities LVA

Mārtiņš Liberts

Three-stage stratified probability proportionate to size (PPS) designs

Notation

d = 27 630 = total number of dwellings to be sampled (main sample and reserve sample together);

D = 683 575 = total number of dwellings in the sampling frame;

$\tilde{d}$ = 15 = planned number of dwellings to be sampled in PSU (main sample and reserve sample together);

d_{hi} = number of dwellings to be sampled in PSU i of stratum h;

$D_{hi} = MOS_{hi}$ = number of dwellings in the sampling frame of PSU i of stratum h, this is also measure of size for PSU;

m_h = number of PSUs to be sampled in stratum h;

n_{hik} = number of persons to be sampled from dwelling k in PSU i of stratum h;

N_{hik} = total number of eligible persons in dwelling k in PSU i of stratum h.

Let $r = d/D$ to be the owerall sampling fraction for dwellings.

PSU selection

Certainty PSUs are those PSUs if $D_{hi} > \frac{\tilde{d}}{r} = \frac{\tilde{d}}{d} \times D = \frac{D}{\tilde{m}}$ where $\tilde{m}$ is the planned number of sampled PSUs.

The probability of selecting PSU i in stratum h is

$$P_{hi} = \begin{cases} \frac{m_h \times D_{hi}}{\sum_{j \in h} D_{hj}} & \text{if} \quad D_{hi} < \frac{\tilde{d}}{r} \\ 1 & \text{if} \quad D_{hi} > \frac{\tilde{d}}{r} \end{cases}$$

DU selection

The *conditional* probability of selecting dwelling k from PSU i in stratum h is

$$P_{k|hi} = \frac{d_{hi}}{D_{hi}}.$$

The *overall* probability of selecting dwelling k in PSU i of stratum h is

$$P_{hik} = \begin{cases} \frac{m_h \times D_{hi}}{\sum_{j \in h} D_{hj}} \times \frac{d_{hi}}{D_{hi}} = \frac{m_h \times \tilde{d}}{\sum_{j \in h} D_{hj}} \approx r & \text{if} \quad D_{hi} < \frac{\tilde{d}}{r} \\ \frac{d_{hi}}{D_{hi}} \approx r & \text{if} \quad D_{hi} > \frac{\tilde{d}}{r} \end{cases}$$

The DU sample size in a PSU is

$$d_{hi} = \begin{cases} \tilde{d} & \text{if} \quad D_{hi} < \frac{\tilde{d}}{r} \\ [D_{hi} \times r]_3 \geqslant \tilde{d} & \text{if} \quad D_{hi} > \frac{\tilde{d}}{r} \end{cases}$$

Please note d_{hi} is rounded to the closest multiple of 3 (because reserve sample is 50 % of the main sample) while preserving the total sample size, namely, $\sum d_{hi} = d$.

Person selection

The *conditional* probability of selecting person l from dwelling k in PSU i of stratum h is

$$P_{l|hik} = \frac{n_{hik}}{N_{hik}}.$$

The *overall* probability of selecting person l from dwelling k in PSU i of stratum h is

$$P_{hikl} = P_{hi} \times P_{k|hi} \times P_{l|hik} = \begin{cases} \frac{m_h \times \tilde{d}}{\sum_{j \in h} D_{hj}} \times \frac{n_{hik}}{N_{hik}} & \text{if} \quad D_{hi} < \frac{\tilde{d}}{r} \\ \frac{d_{hi}}{D_{hi}} \times \frac{n_{hik}}{N_{hik}} & \text{if} \quad D_{hi} > \frac{\tilde{d}}{r} \end{cases}$$

or we can write

$$P_{hikl} \approx r \times \frac{n_{hik}}{N_{hik}}.$$