Estimation in a mixed-mode, web and face-to-face, survey

Kaur Lumiste
University of Tartu
kaur.lumiste@ut.ee

Euroopa Liit
Euroopa Sotsiaalfond

Eesti tuleviku heaks

BNU Network on Survey Statistics
August 24th 2012 Valmiera

Introduction

- In 2012 4 countries in the European Social Survey (ESS) network will have a mixed-mode experiment (Estonia, Slovenia, Sweden and the UK)

- The aim of the experiment is to test for mode effects and develop means and protocols to avoid them.

Introduction

- Web and telephone survey modes are used in conjunction with the traditional face-to-face interview
- The ESS in Estonia will experiment with a web and face-to-face mixed-mode survey design
- Current paper is a short presentation of preliminary results of estimating study variables in this experimental design

Preliminaries

Preliminaries

Notations

- Finite population: $U = (1, 2, \ldots, N)$
- Design vector: $\mathbf{I} = (I_1, I_2, \ldots, I_N)$
- Inclusion probability: $E(I_i) = \pi_i > 0$
- Design weigths: $a_i = 1/\pi_i$

A sample *s* of dimension *n* is selected from *U.*

In case of non-response data can only be collected from a sample subgroup $r \subseteq s$.

Preliminaries

Some more notations:

- Study variabel: y
- Population total: $Y = \sum_U y_i$
- Auxiliary variable: $\mathbf{x}_i : J \times 1$

Basic design unbiased estimator:

$$\hat{t}_{\mathrm{HT}} = \sum_s a_i y_i$$

Estimation in the mixed-mode survey

Estimation in the mixed-mode survey

Since the data will be collected in two parts, we can define two random vetors:

$$\mathbf{W} = (W_1, W_2, \ldots, W_n \,|s)$$
$$\mathbf{F} = (F_1, F_2, \ldots, F_n \,|s)$$

Where $W_i = 1$ if unit *i* answered in the web mode and is 0 otherwise and $F_i = 1$ if unit i answered in the face-to-face mode and is 0 otherwise.

Estimation in the mixed-mode survey

As the data is collected, the sample is first divided into two subgroups:

$$r_{web} = \{i \mid W_i = 1, s\}$$

and

$$s_{ftf} = s - r_{web} = \{i \mid W_i = 0, s\}$$

We can now define $P(W_i = 1 \mid s) = p_i$ and an unbiased estimator:

$$\hat{t}_{web} = \sum_{r_{web}} \frac{y_i}{\pi_i p_i}$$

Estimation in the mixed-mode survey

Sampled persons, who do not answer in the web survey are approached for a F2F interview, the probability of that happening is 1-p_i .

By defining $P(F_i = 1 \mid s_{ftf}) = q_i$ we can construct another unbiased estimator:

$$\hat{t}_{ftf} = \sum_{r_{ftf}} \frac{y_i}{\pi_i (1 - p_i) q_i}.$$

Now

$$\hat{t} = \alpha \hat{t}_{web} + (1 - \alpha) \hat{t}_{ftf}$$

Estimating mode participation probabilities

Estimating mode participation probabilities

Särndal (2011) uses auxiliary information to estimate response probabilities $\theta_i = P(i \in r \mid s)$.

For estimating θ_i we need two conditions:

1. $\hat{\theta}_i = \boldsymbol{\lambda}'\mathbf{x}_i.$
2. $\sum_s a_i(I_i - \hat{\theta}_i)\mathbf{x}_i = 0$ or $\sum_r a_i\mathbf{x}_i = \sum_s a_i\hat{\theta}_i\mathbf{x}_i.$

Using these conditions we can estimate

$$\hat{\theta}_i = \left(\sum_r a_i\,\mathbf{x}_i\right)'\left(\sum_s a_i\mathbf{x}_i\mathbf{x}_i{}'\right)^{-1}\mathbf{x}_i.$$

Estimating mode participation probabilities

From this result we can estimate the mode participation probabilities:

$$\hat{p}_i = \left(\sum_{r_{web}} a_i \mathbf{x}_i\right)' \left(\sum_{s} a_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \mathbf{x}_i$$

and

$$\hat{q}_i = \left(\sum_{r_{ftf}} a_i \mathbf{x}_i\right)' \left(\sum_{s_{ftf}} a_i \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \mathbf{x}_i.$$

Future research

Future research

- Properties of $\hat{t}_{web}$ and $\hat{t}_{ftf}$
- The optimal α if $\hat{t}_{web}$ and $\hat{t}_{ftf}$ are dependant
- Properties of mode participation probabilities
- Improving the initial estimators (calibration?)
- A simulation study
- ....

Thank you for your attention!