Mixed-mode Census 2021 survey with voluntary part in Estonia to improve quality of the estimates

Kristi Lehto[1]

[1]*Statistics Estonia, Estonia*

Abstract

The 2021 census in Estonia was mostly based on administrative data. All EU-mandatory characteristics were collected from administrative data. However, the purpose of the sample survey was to collect information on persons living in Estonia that is not available in the registers (religious affiliation, knowledge of languages and dialects, existence of a long-term illness or health problem and health-related limitations on daily activities).

In 2011 census, when Estonia used the first time of self-enumeration (CAWI - Computer Assisted Web Interview) the response rate of CAWI was 67%. This gave an idea to use the voluntary CAWI part in 2021 census survey.

The probability sample included approximately 40,000 dwellings (around 30,000 of these inhabited), i.e. around 60,000 persons for whom participation in the population and housing census was mandatory according to the law. In CAWI mode all those who wished could respond voluntarily even outside the sample. CAWI response rate was 43,1% which is very high concerning that answering was voluntary.

CAWI respondents are different from CATI/CAPI (Computer Assisted Telephone Interview/ Computer-Assisted Personal Interview) respondents. They are younger, healthier, less religious and know more foreign languages based on Census 2011. In order to obtain the unbiased estimates, it is necessary to skillfully combine the data of different modes. The estimate extends the proportion of the surveyed characteristic found in CAWI respondents to CAWI population, and the proportion found in CATI/CAPI respondents to the rest of population.

To mix probability and non-probability (voluntary) sample helped to improve quality of the census estimates and publish more detailed breakdowns.

Keywords: census survey, mixed-mode, survey with voluntary part

1. Introduction

Counting only EU-mandatory characteristics one could say that Population and Housing Census 2021 was register based Census in Estonia. For using registers, Statistics Estonia has carried out methodological and information technology preparatory work since 2010. (Statistics Estonia, 2022 and Statistics Estonia, 2022 June).

There was a need for the sample survey to collect information on persons living in Estonia that is not available in the registers (religious affiliation, knowledge of languages and dialects, existence of a long-term illness or health problem and health-related limitations on daily activities). These census variables were also collected in 2011 when there was mostly traditional census in Estonia and they are mentioned in Statistics Act. Instead of asking from everyone, the sample survey was proposed.

2. Census survey

The great advantage of the sample survey is to calculate the accuracy of the estimates according to the survey design. Even in censuses there are some non-response and it is hard to evaluate the influence of non-response to the output variable.

Sample design for Census survey was developed in Statistics Estonia in cooperation with Tartu University. Survey design was tested on the Census 2011 data. More information about Census survey is on the Statistics Estonia website (Statistics Estonia, 2022 November).

2.1 Sample design

Starting point for the sample design was the output, as all results were necessary to publish for local government units (including Tallinn districts) level, which are very different in size in Estonia, so stratified sample was the obvious choice.

The next point was survey mode, it was clear that the first mode will be CAWI mode as 2011 Census it worked very well and followed by CATI/CAPI mode. As in last census CAWI response rate was 67% there were an idea do not to use CAWI for only sample persons but all those who wish could respond voluntarily even outside the sample. To suport this idea, a media campaign was made.

Using all this information the final sample design was:

- Stratified systematic sampling from dwellings
- Stratification is made by local government units

Why sample for dwellings? Dwellings are always in same place and easier to find instead of people who can move around.

We tried many different sample sizes in local government units and realized that this isn't very useful, thus final sample size is in table 1. Census survey sample size was much bigger than any other Statistics Estonia surveys due to the obligation to publish results on local government unit level. In other surveys the lowest level to publish estimates is county level.

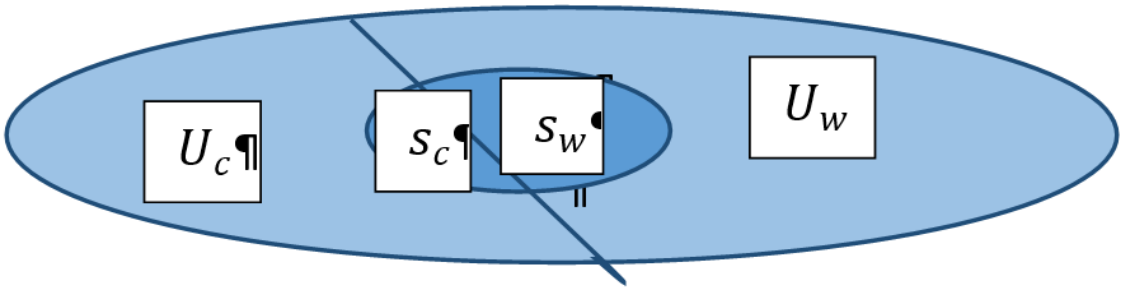
Table 1: Stratification and sample size of Census survey

Size of municipality/Tallinn district	Size of sample
Municipalities with less than 2,001 residents, i.e. small islands	all residents
Municipalities with 2,001 to 20,000 residents	300 inhabited addresses
Municipalities with 20,001 to 49,999 residents	500 inhabited addresses
Municipalities with 50,000 to 100,000 residents (incl. Tartu)	800 inhabited addresses
Tallinn districts	500 inhabited addresses
Total – 78 municipalities and 8 Tallinn districts	**approximately 40,000 dwellings (around 30,000 of these inhabited)**

2.2 Weighting

For weighting, it was important to keep in mind that CAWI respondents are different from CATI/CAPI respondents. They are younger, healthier, less religious and know more foreign languages based on Census 2011. In order to obtain the unbiased estimates, it is necessary to skillfully combine the data of different modes (see figure 1).

Figure 1: Target population and sample by mode



Where $U = U_w \cup U_c$ – target population, $w = web = CAWI$, $c = CATI/CAPI$

$s = s_w \cup s_c$ – simple random stratified sample from *U*

$U_w - s_w$ – CAWI non-probability (voluntary) part

All households with at least one CAWI respondent belongs to the CAWI population, the rest of the people belong to the CATI/CAPI population. Households that were originally included in the random sample, but responded via CAWI, are part of the general population of CAWI. Also four small islands are included in the CAWI population, where the collection took place everywhere, but responses from all people was not received. The population of the CATI/CAPI

is described by people who were randomly sampled and answered to the interviewers (CATI/CAPI).

The calculation of the design weights and the calibration of the weights were done separately in two parts of the population. The design weight is the number of people in the local government unit divided by the number of respondents in the local government unit. The CAWI part has a design weight of approximately 1, the CATI/CAPI part the design weight is on average 40. Calibration of the design weights for background characteristics of the population (e.g. age, gender) to known amounts was necessary because the response was not the same in all groups.

The distributions of the background characteristics of both parts of the population are known from the register-based census. For calibration the gender and age distribution were used at the local government level, and education, ethnicity, degree of urbanization "status in the household" and "status in the family" divisions in the whole Estonia level.

Using background characteristics to calibrate the weights ensures that respondents are as representative of the entire population as possible, but it should be kept in mind that the purpose of the weighting is to calculate estimates and standard errors to the questions asked.

Since the calibration is done separately for CAWI and CATI/CAPI part, we can summarize the estimates

$$\hat{t}_{1cal} = \hat{t}_{CAWI,cal} + \hat{t}_{CAPI,cal}.$$

And the variances can also be summarized, thus the standard error is expressed as follows

$$\text{sd\_}cal = sd(\hat{t}_{1cal)} = \sqrt{[sd(\hat{t}_{CAWI,cal})]^2 + [sd(\hat{t}_{CAPI,cal})]^2}.$$

3. Results

CAWI response rate inclunding voluntary part in Census survey was from 20,3% in Narva city to the 61,2% in Kiili municipality. For whole Estonia CAWI response rate was 43,1%.

Standard errors for Census survey main indicators in local government unit level are published in Statistics Estonia website (Statistics Estonia, 2022 November). Overall expectation was that no standard error for the share is not bigger than 5%. This goal was achieved. The biggest standard error in local government unit level for the share of the affiliated persons, for the share of personas who can speak at least one foreign language and for the share of person who have a long term illness or health problem was accordingly 4.1, 4.3 and 2.6.

To estimate the influence of voluntary part to the accuracy we calculated second weights to take into account only persons who belonged to the initial probability sample (excluding voluntary part).

Standard errors were calculated for using both weights. As expected, standard errors were bigger when using only initial sample in most local government units (see table 2). The biggest effect of the voluntary part was to the accuracy of the share of persons who have a long term illness or health problem. There were only 3 local government unit where the voluntary part didn't improve the accuracy. For the religiously affiliated and foreign language estimates the accuracy didn't improve in 10 and 11 municipalities accordingly. In all other local government units the standard error were smaller using the voluntary part of the Census survey.

Table 2: Simple random sample vs simple random sample with voluntary part – standard errors in the local government unit level

Criteria	The share of the religiously affiliated persons, at least 15 years old	The share of persons who can speak at least one foreign language, at least 3 years old	Share of persons who have a long-term illness or health problem
Standard error is bigger in simple random sample	76	75	83
Standard error is smaller in simple random sample	10	11	3
Number of municipalities/Tallinn districts	**86**	**86**	**86**

There was no concrete connection between the CAWI response rate and the accuracy. That mean the higher CAWI response rate doesn't necessarily guarantee a lower standard error, because of the variance of the estimated variable is not related to the CAWI response rate in the local government unit level.

4. Conclusions

The aim of this paper is to compare the accuracy using voluntary part in Census survey and simple random sample.

The large number of volunteers in the Census survey helped to improve the accuracy of the estimates compared to the simple random sample. The standard error was smaller when using voluntary part of the survey in most local government units.

Improvement of the accuracy wasn't connected to the voluntary rate. It means that voluntary part may not always help to improve the accuracy of the estimates.

More important than an accuracy is the opportunity for the users to get the estimates in more detailed level, for this purpose voluntary part in Census survey was a huge help. Detailed breakdowns belong to the Census output and was one of the reasons to use voluntary CAWI part in Estonia Census survey 2021.

Voluntary part didn't add the extra cost to the Census survey, it was only demanding to the methodologist to take into account every aspects in weighting process.

References

Statistics Estonia (2022) *Population and Housing Census. Methodology* https://rahvaloendus.ee/en/census-2021/methodology

Statistics Estonia (2022, June) *Description of the 2021 register-based census methodology* https://www.stat.ee/sites/default/files/2022-06/Registrip%C3%B5hise%20loenduse%20metoodika%20raport_EN.pdf

Statistics Estonia (2022, November) *Description of the sample survey methodology for the 2021 census* https://www.stat.ee/sites/default/files/2022-11/Loenduse%20valikuuringu%20metoodika%20raport.pdf (only in Estonia)