Distribution of surveys sample in Central Statistical Bureau of Latvia.

Andris Fisenko[1]

[1]Central Statistical Bureau of Latvia, e-mail: andris.fisenko@csb.gov.lv

Abstract

The goal of this paper is to show sampling strategy for Labour force surveys in Central Statistical Bureau of Latvia (CSB). *Keywords*: Social survey, LFS, rotation pattern, distribution.

1 Introduction

LFS sampling design is sampling design with following features:
- Rotation of units according to the specific rotation pattern
- Uniform distribution of sampled units over space
- Uniform distribution of sampled units over time
- Easy management of sampled units in a sample
- Easy coordinate with different continuous social samples.

The paper describe detailed about sampling frame creation and distribution of surveys sample.

2 Sampling frame

Two sampling frames are built for each sampling stage due to two-stage sampling is used.

At the first stage the list of sampling areas is used as sampling frame. The sampling areas are originally created from the list of population census counting areas 2000 and revised according to the Population register data in 2009. The main reason for redesign PSU was that the size of PSUs is outdated. The largest differences were found in rural areas and capital city – Riga. For coordinated social survey is important to provide PSU size with enough dwellings for all surveys. Other reason was that new technology solutions where introduction with available auxiliary information. There are several administrative data source is used to create this areas. Putting together Population register and Building register which are used as a basis for future activities. Almost for each dwelling we have geographical coordinates and this information allow to redesigning PSU size. Calculation design weights it's important that PSU are similar as possible by size (number of dwellings) to avoid big variance and each dwelling have equal including probability. Latvian territory is covered by ~4000 sampling areas (PSU

At the second stage sampling frame is built from the Demographic Statistics Data Processing System. The Demographic Statistics Data Processing System includes data from Population register, The National Real Estate Cadastre Information System etc. Each sampling area contains ~500 dwellings.

3 Sampling design

3.1 Primary sampling Unit (PSU)

The last sampling redesign was made in 2009 and it was introduced in practice since 2010 and is used till now. The redesign of the sampling design and the selection of a new PSU sample were started after the update of the PSU frame. The resulting sampling design is a probability two-stage sample design. Areas are the first stage units. Areas are sampled by stratified systematic sampling with sampling probabilities of areas being proportional to the size of areas (measured by the number of dwellings). The stratification of areas is done by the level of urbanization. There are four strata: Riga (the capital of Latvia), other cities, towns and rural areas. The sampling is done with random starting point in each stratum. Each group of PSU is numerated. The PSU are listed by geographical region, and within a geographical region in a serpentine order that places units containing similar types of people together.

Image 1 show how is realised serpentine order in rural area.

Image 1: Serpentine over Latvia
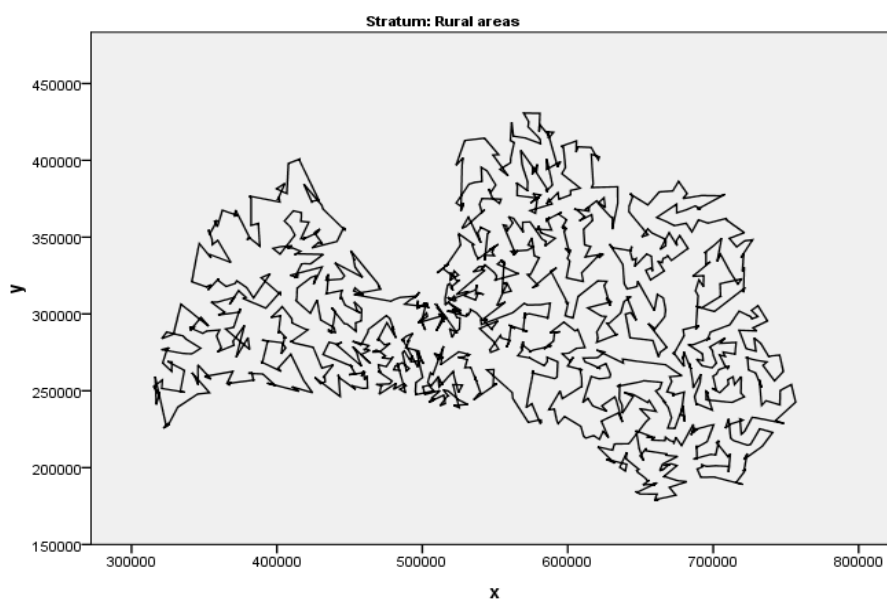
The design allows organising rotating panel surveys. This is an important requirement for LFS. The Latvian LFS is a rotating panel survey with rotation scheme 2-(2)-2 (European Commission, 2012b, 7.p). The design provided easy management of sampled PSUs in practice. The sample of areas can be generated for several years – allowing the timely planning of workload for interviewers. It is not necessary to make any corrections to the sample of areas during the usage of it. The sampling design is suitable for application of variance estimation methods based on re-sampling, for example, dependent random groups or Jackknife (Wolter, 2007). A new sample of areas was selected after the redesign of sampling design. The area sample selected can be used

simultaneously for several sample surveys. The coordination of samples for three social surveys (LFS, HBS, SDT) is incorporated in the design. The coordination of three samples allows achieving lower total fieldwork cost comparing to non-coordinated samples. The sample of areas is usable also for one-time surveys. For example, the selected area sample has been used in Latvia for European Health and Social Integration Survey (EHSIS) in 2012. The area sample can be used for household sampling (for example, LFS, HBS, SDT) or individual sampling (for example, EHSIS).

For better visualization all PSU that are numerated in serpentine order we put on the circle. Each dot in image 2 represented 1 PSU in a 1 week and quarter.

Image 2: Sample of 1 week and 1 quarter.
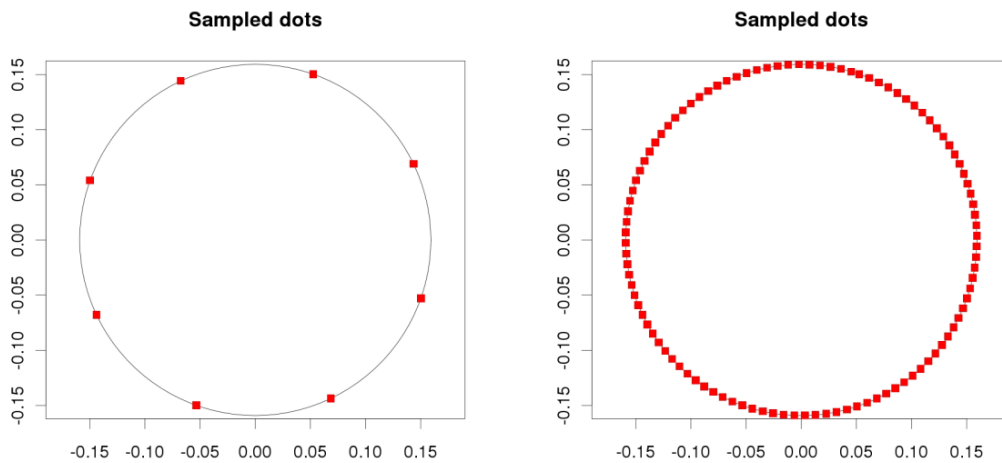
Since LFS use 4 wave rotating panel surveys, not in all PSU are selected HBS and SDT sample. The PSU samples of HBS and SDT are sub-samples of LFS PSU sample. Table 1 show how are samples distributed based on LFS.

Table 1: Samples for three social surveys.

Survey	1.wave	2.wave	3.wave	4.wave
LFS				
HBS	✕			
SDT	✕			

Each sampled PSU are used for 3 surveys (except 1.wave of LFS) in this connected survey situation. Each PSU is included in the sample for 8 consecutive quarters, and then it is out of the sample for a long time. The reduction of the total travel costs is one of important aspects what are taken into account when planning a survey.

Simple example shown in image 3 where we can see differences between two sample

designs and travel distance for one interviewer.

Image 3: Two stage and simple random sample.
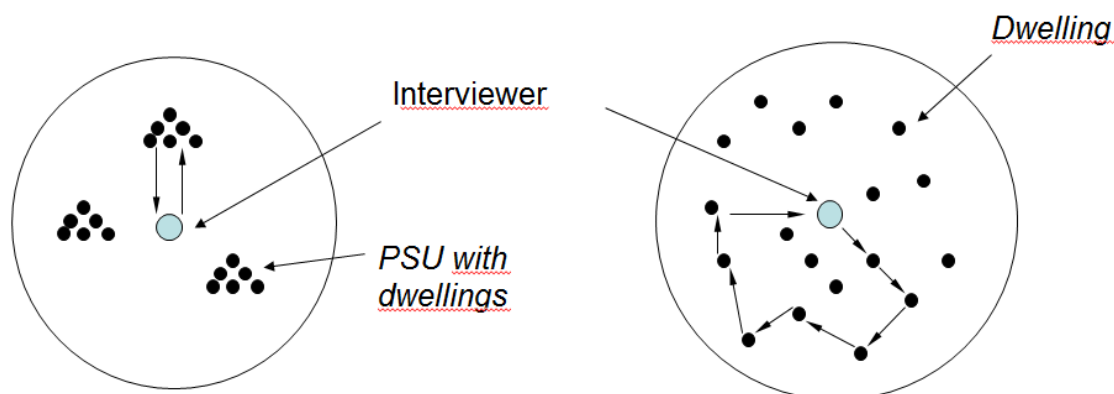
In this case there are possibilities to coordinate different continuous household samples in PSU. The interviewer can manage to do all three surveys in a PSU with low travel times. This decision reduces the travel costs of interviewer and total costs for all three surveys. More about cost function for survey can be read in Mārtiņš Liberts Doctoral Dissertation (Liberts, 2013).

3.2 Secondary sampling unit (SSU)

At the second stage within each sampled PSU dwellings are selected as non-coordination by a simple random sampling for each survey (LFS, HBS, and SDT).

There are order in which sequence are samples of three social surveys are sampled. LFS is first sampled, then HBS and last SDT. This is done in reason what LFS is biggest survey and used more PSU than HBS or SDT.

4. Conclusions

We have found a way how to compute sampling units for LFS and other continuous surveys according to the rotation pattern. Primary sampling units can be computed for long period (5 years for example). It allows timely planning of the work for interviewers. Two stage sampling design is more effective comparing to single stage sampling design – efficiency measured by cost and precision for three coordinated social surveys. It is possible to use the PSU sample selected by the design for other social sample surveys.

References

Lapiņš,J., Vaskis E., Priede Z. & Bāliņa S. (2002). *Houshold surveys in Latvia.* Central Statistical Bureau of Latvia, Riga.
Mārtiņš Liberts "The Optimisation of sampling design." University of Latvia, 2013
Wolter, K. M. (2007). Introduction to variance estimation. New-York:
Springer.