

Peteris Vegis (Central Statistical Bureau of Republic of Latvia)

Maranda Behmane (Central Statistical Bureau of Republic of Latvia)

Possibilities to use administrative data sources for register based Census in Latvia

Topic 1 – Bringing in information from where we can get it

Keywords: population, employment, unemployment, activity, occupation

Introduction

The 2021 Population and Housing Census (2021 Census) in Latvia will be carried out on base of information obtained from administrative data sources, as well as regular statistical surveys. Work aims on studies of administrative data availability, assessment of the quality of the incoming administrative data, development of administrative data processing methodologies and assessment of the quality of data processing.

To determine the impact of the quality of administrative data on statistical quality such important quality indicators as timeliness, relevance, comparability with other data sources, and others should be assessed. On the other hand, administrative data processing methodology should be evaluated in data processing areas such as data editing, detection and treatment of „outliers”, imputation necessity and others.

The first main indicator is the number of usual residents. Work on estimation method was started in 2012. From October 2015 to September 2016 CSB of Latvia has an Eurostat grant project "Improvement of the use of administrative sources (ESS.VIP ADMIN WP6 Pilot studies and applications)".

The main objective of the project is to make initial evaluation of the administrative data availability and quality, to understand and identify problems of obtaining economic characteristics from administrative data sources, how they could be linked together in the Data Warehouse for Social Statistics (DWHSS) and to start work on estimations and imputations

Methods / Problem statement

The 2011 Population and Housing Census (2011 Census) showed a difference of 7% between Census data and The administrative Population Register (PR) of the Ministry of Interior. It was decided in 2012 to work out a new method for estimating the annual number of population. The method is based on the PR and other administrative register data.

The aim of the method is to estimate the status of the actual place of residence at the beginning of the year for every resident registered (on individual level) in Latvia. The Logistic regression model had been developed to evaluate the residence status. The Grant project includes feasibility study on use of administrative data and information from statistical sample surveys to obtain data for the 2021 Census core topics on economic characteristics of population - current activity status, occupation, industry (branch of economic activity) and status in employment.

As it is planned to use administrative data sources not only for the 2021 Census, but for other statistical areas, working out of the administrative data quality system was initiated in Latvia in 2016. Dutch experience is used as a model.

Results / Proposed solution

With the help of administrative register data on each person registered within The PR there are more than 200 characteristic variables developed. Annual improvements of the method are done. At the beginning of 2016 43 administrative data sources are used to estimate the population figures. This year their number will be increased with information on a person's level on: students studying abroad, visitors of libraries, personalized e-ticket users in Riga, seamen and diplomats.

As result of the Grant project administrative data sources containing information on the economic activity of population will be identified and description which economical characteristics of population needed for the 2021 Census can be derived from which particular administrative data base available for the CSB of Latvia. Data will be incorporated into the DWHSS. Administrative data processing methodology will be worked out and described. The "outliers" values will be detected and proposals for further steps to work out administrative data processing methodology as regards economic characteristics of population will be worked out.

Preparation of the list of conditions to be met will be worked out and proposals for future work on imputation methodology to be used for economic characteristics of population included in the 2021 Census will be prepared.

Conclusions

Estimation method worked out by the CSB of Latvia will help to have the 2021 Census population number for Latvia more precise as it would be in case of direct use of the PR. Recalculation of the number of population for years between two Censuses (as it was done before) will not be needed. Work done within the Grant project will be a good base to obtain high quality economical characteristics of the population of Latvia for the 2021 Census. Results of the Grant project will be used for other areas of Social Statistics and as an input for working out of the administrative data quality system within the CSB of Latvia.

The next planned step (probably as another Grant project) is to make initial evaluation of the administrative data availability and quality, to understand and identify what are problems of obtaining educational characteristics of the population from administrative data sources, how they could be linked together in the DWHSS, to introduce the ISCED 2011 and to start work on estimations and imputations methodology.