Defining the population size using the residency index. Case of Estonia

Mare Vähi[1], Ethel Maasing[2] and Ene-Margit Tiit[3]

[1] *University of Tartu, Estonia, mare.vahi@ut.ee*
[2] *Statistics Estonia, Estonia, ethel.maasing@stat.ee*
[3] *University of Tartu and Statistics Estonia, Estonia, enemargit.tiit@stat.ee*

Keywords: residency index, discriminant analysis

Census data have been the most important and valuable data in population statistics from the very beginning of scientific thinking. But nowadays, when there exist different information sources, the reliability and exactness of census data does not satisfy. Here are two reasons: probably, the coverage of census has fallen due to mobility of people. Also, needs of researchers are higher today. After the Estonian census 2011 the census team found that there was some under-coverage present. To determine the amount of non-enumerated people the following procedure was used. The set of people belonging to Estonian population register as residents, but not enumerated in census 2011 were regarded as potential residents. All existing administrative registers were used to define the signs of life for these people: activity in a register during 2011 gave to a person a sign of life. The signs of life were used as binary variables to discriminant the residents and non-residents using several multivariate technics (linear and logistic discriminant analysis). Training groups consisted of the "true residents" (who belonged to PR as Estonian residents and were enumerated) and "true non-residents" (who were in PR as residents of foreign countries and also not enumerated as people living currently in Estonia).

As a result, the under-coverage of about 2,3% was found. The error of decisions was not more than 5%. The following task was to use the methodology for following years and to cover the whole population. Here the following problems arose: 1) No training groups exist for subsequent years; 2) Making decision by years the continuity and stability of population might disappear. Hence we decided to define for each person from the population a residency index between 0 and 1 that will be recalculated annually using the signs of life. If the value of index of a person is higher than threshold, then he/she is resident, if it drops below the threshold, the person is excluded from the set of residents. The recalculation formula of residency index $R(i, j)$ for person i in year j is the following:

$$R(i, j) = d(Ri, j - 1) + g \sum a_k X(i, j - 1, k),$$

where $R(i, j-1)$ is the index of the person in last year, $X(i, j, k)$ is the value of k-th sign of life of the person i in year j and a_k is the weight of the sign of life k. The parameters d and g have the value between 0 and 1. If the value of $R(i, j)$ is bigger than 1, it will be truncated to 1. All persons having the value of R higher than threshold b are considered as residents in year j, the others are non-residents, but will stay in the population and will be able to get the status of resident in future. The values of parameters d, g and b are defined from some theoretical concepts connected with the acceptable time of change the status. The weights a_k have been defined in different ways: all equal to 1 (simple sum of signs of life), proportional to their description value and also using some logarithmic scale. The efficiency of the index has been tested using a series of years: 2012, 2013, 2014 and 2015. The comparison of estimated number of residents and the official number of residents calculated by traditional way in Statistics Estonia has been made.