

Post Enumeration Survey (PES) of the 2011 Population Census

Methodological Note

1. Introduction

The population census aims at the full listing of persons residing in the country at a given time. However, the experience of former censuses has shown that this is not always achieved with absolute precision, despite the intensive efforts of all parties. This is possibly due to the large-scale of the census tasks.

Errors in censuses can appear due to several reasons such as mistakes in the questionnaires, insufficient written instructions, unsuccessful training of the field workers, unclear procedures ect. The errors mentioned above coming from the census design, are known as measurement errors.

Another type of error, is coverage error which refers to the undercount of survey units due to omissions or the overcount of survey units due to duplications and erroneous inclusions. In other words there are three types of coverage error: a) omissions, b) duplications, and c) erroneous inclusions.

In addition, content error is the error found in the characteristics of the enumerated survey units eg. male that is characterised as mother in family relations.

The PES is a sample survey conducted in order to assess the quality of the results of the census, regarding:

- (a) the undercount or overcount of the total resident population. It is possible to design the PES in such a way that reliable estimates of undercount or overcount are obtained for geographical areas.
- b) the characteristics of the population, by measuring the level of agreement for responses to questions on selected characteristics (age, sex, marital status, occupational status ect.) among census and PES.

The study focus of the PES is the Enumeration Sections which are the smallest census units. The dwellings of the Enumeration Sections can be divided into two categories: a) dwellings that were not possible to get listed by the enumerator and b) dwellings that were enumerated. In the first case (a), it is obvious that all the household members who were staying in the dwelling were not enumerated. However, even in the second case (b), it is possible to observe errors, such as one or more members of the household may not have been enumerated.

The PES of the 2011 Population Census was conducted during the period 15 to 30 June 2011.

2. Sample design

The PES is a sample survey with final survey unit the household. The multi-stage stratified sampling was applied for the selection of the sample. The sample size was about 19.380 dwellings and the total sampling fraction was about 0,32%. The size of the sample was determined in order to estimate the coverage and the quality of the census data in a lower administrative level than the whole country and in particular at the Municipal level (Kallikratis plan).

3. Stratification

The first criterion of stratification is the geographical division of the country. In each Municipality (NUTS 4), the stratification is conducted by allocating the Municipal / Local Communes according to the degree of urbanization (urban, semi-urban, and rural regions). Except for the two former Major City Agglomerations (Athens and Thessaloniki), the produced strata according to the degree of urbanization are:

Urban	Stratum	1	Municipal Communes with 10.000 inhabitants or more
Semi-urban	"	2	Municipal and Local Communes with 2.000 to 9.999 inhabitants
Rural	"	3	Local Communes up to 1.999 inhabitants

In consequence, the households are stratified at the final stratum, which is made up from the crossing of the stratification criteria.

The former Greater Athens Area is divided into 31 strata of about equal size (equal number of households) on the basis of the lists of city blocks of the Municipalities that constitute it and taking into consideration socio-economic criteria. Similarly, the former Greater Thessaloniki Area is divided into 9 equally sized strata. The two former Major City Agglomerations account for 40% of the total population.

4. Sample selection of dwellings

For the sample selection of the dwellings under survey the three stage sampling was applied, with primary unit the Municipal / Local Commune, secondary unit the Enumeration Section, tertiary unit the dwelling and final sampling unit the household. In the strata where the number of Municipal / Local Communes was not over two, the two stage sampling was applied, with primary unit the Enumeration Section and secondary unit the dwelling, with final survey unit the household.

The three stages of sampling are described analytically as follows:

1st stage of sampling: In each stratum at least two primary units (Municipal/ Local Communes) were selected with replacement and probabilities proportional to their population sizes, according to the census data of 2001. In the cases where the two stage sampling was applied, all the Municipal / Local Communes were included in the sample.

2nd stage of sampling: Selection of at least two Enumeration Sections from each Municipal or Local Commune, with probability proportional to their size based on their number of dwellings, according to the census data 2011. In the cases where this information was not available, the Enumeration Sections were selected with equal probabilities.

3rd stage of sampling: In each Enumeration Section under survey, the first aim of the field worker was to recognize carefully and precisely the limits of the Section from the sketch or the map. In consequence, he/ she should have enlisted in the form DE-1 all the dwellings or other type of buildings (conventional or not) found within the limits of the Section. In other words, the field worker would move from building to building listing the dwellings, either occupied or vacant. As a result, the list included every position that could have been occupied by a household at the day of the census, within the limits of the Section. The form DE-1 was used as sampling frame for the selection of the systematic sample of 10 dwellings, with equal probabilities.

5. Estimation of the usual resident population and of the coverage error

The members of the households of the sample were categorized as follows:

1. People who resided in the same dwelling on the census date and on the PES date (non-movers)
2. People who resided in the dwelling on the census date and left it after the census/ did not reside there on the PES date (out-movers)
3. People who resided in the dwelling at the PES date but at the census date they were not enumerated in this dwelling (in-movers)
4. People who did not belong to the target population on the census date, for example a baby born after the census date or a person residing abroad on the census date but lived in the household on PES date (out of scope)

In every Enumeration Section we matched the members of the sample households with the members of the households of the census. In order to proceed with the matching, the criteria used in several combinations were full name, gender, legal marital status, year of birth, place of birth, citizenship, work status, educational level and the Municipal Unit of registration.

The following figures resulted from the matching procedure:

n_{PES} : number of people captured in the PES

n_{CEN} : number of people captured in the census

n_{MAT} : number of people captured in the census and in the PES
(matched)

n_{ERR} : number of people erroneously included because they are out of the scope of the survey

For the extrapolation of the above figures in a geographical level, firstly they are extrapolated at stratum level (stratum estimation) and next the estimations of all strata are added up composing a certain geographical level. The estimation of each stratum is held by multiplying the number of people of each category mentioned above with the proper extrapolation factor. The extrapolation factor for each person, is the same to the extrapolation factor of the household and is calculated as follows:

Signs:

h : final stratum

N_h : number of Municipal / Local Communes in the stratum h

n_h : sample size of Municipal / Local Communes in the stratum h

X_h : number of households in the stratum h , according to the 2001 census data

X_{hi} : number of households of the selected Municipal / Local Commune of order i
($i=1, 2, \dots, n_h$) of the stratum h , according to the 2001 census data

A_{hi} : number of Enumeration Sections of the Municipal / Local Commune of order i of the stratum h , according to the 2011 census data

a_{hi} : number of Enumeration Sections selected at the Municipal / Local Commune of order i of the stratum h

Y_{hi} : number of dwellings of the Municipal / Local commune of order i of the stratum h , according to the 2011 census data (where available)

Y_{hij} : number of dwellings of the selected Enumeration Section of order j ($j = 1, 2, \dots, a_{hi}$), of the Municipal / Local Commune of order i of the stratum h , according to the 2011 census data (where available)

M_{hij} : number of dwellings of the form DE-1 which was used as sample frame for the selection of the systematic sample of dwellings, in the Enumeration Section of order j , of the Municipal / Local Commune of order i of the stratum h

m_{hij} : dwellings sample size in the Enumeration Section of order j , of the Municipal / Local Commune of order i of the stratum h

Extrapolation factors

1. In the case of applying three stage sampling, where there is available data for the number of dwellings in the Municipal / Local Communes of the sample, the extrapolation factor w_{hij} is given by the formula:

$$w_{hij} = \frac{1}{n_h} \cdot \frac{X_h}{X_{hi}} \cdot \frac{Y_{hi}}{a_{hi} Y_{hij}} \cdot \frac{M_{hij}}{m_{hij}} .$$

2. In the case of applying three stage sampling, where there is no available data for the number of dwellings of the Municipal / Local Communes of the sample, the extrapolation factor w_{hij} is given by the formula:

$$w_{hij} = \frac{1}{n_h} \cdot \frac{X_h}{X_{hi}} \cdot \frac{A_{hi}}{a_{hi}} \cdot \frac{M_{hij}}{m_{hij}} .$$

3. In the case of applying two stage sampling, the above formulas are reformed respectively as follows:

$$w_{hj} = \frac{Y_h}{a_h Y_{hj}} \cdot \frac{M_{hj}}{m_{hj}}$$

and

$$w_{hj} = \frac{A_h}{a_h} \cdot \frac{M_{hj}}{m_{hj}}$$

where:

A_h : number of Enumeration Sections of the stratum h , according to Population Census 2011

a_h : number of Enumeration Sections selected in the stratum h

Y_h : number of dwellings of the stratum h , according to the Population Census 2011 (where there was available data)

Y_{hj} : number of dwellings of the selected Enumeration Section of order j ($j = 1, 2, \dots, a_h$), of the stratum h , according to the Population Census 2011 (where this data were available)

M_{hj} : number of dwellings of the form DE-1 that was used as sample frame for the selection of the systematic sample of dwellings, in the Enumeration Section of order j , of the stratum h

m_{hj} : sample size of dwellings in the Enumeration Section of order j , of the stratum h

Consequently, in any geographical level the following estimations can be calculated:

\hat{N}_{PES} : Estimation of PES population

\hat{N}_{CEN} : Estimation of census population

\hat{N}_{MAT} : Estimation of matched persons

\hat{N}_{ERR} : Estimation of erroneously included in the census population

\hat{N} : Estimation of the usual resident population

With the implementation of the method «Dual System Estimation», the estimation of the usual resident population is given by the formula:

$$\hat{N} = \frac{\hat{N}_{PES} \left(\hat{N}_{CEN} - \hat{N}_{ERR} \right)}{\hat{N}_{MAT}}$$

The net coverage error as the rate of people missed in the census, is given by the formula:

$$\text{Net coverage error (\%)} = \frac{\hat{N} - \hat{N}_{CEN}}{\hat{N}} \cdot 100$$

6. References

Cochran, W.G. (1977). Sampling Techniques, New York: John Wiley and Sons

Kish, L, (1965). Survey Sampling, New York: John Wiley and Sons

Kish, L., (1987). Statistical Design and Research, New York: John Wiley and Sons

World Population and Housing Census Programme, (2010). Post Enumeration Surveys, Operational guidelines, Technical Report, New York: United Nations Secretariat, Department of Economic and Social Affairs, Statistics Division