

# **THE METHODOLOGY OF THE 2011 NATIONAL POPULATION AND HOUSING CENSUS – selected aspects**

## **I. GENERAL COMMENTS**

### **1. Legal basis, deadline and the scope of the 2011 population and housing census**

The 2011 Polish Census of Population and Housing (NSP 2011) was the first census performed since Poland's accession to the European Union and took place on the territory of the Republic of Poland **in the period from 1 April to 30 June 2011 as of 31 March 2011, at 00.00.**

The frames of the 2011 population and housing census subject matters, its scope, form, mode, statistical obligations limits and freedom of participation in surveys were provided for in the Act of 4 March 2010 on Polish population and housing census in 2011 (Journal of Laws of 26 March 2010, No 47, item 277) together with implementing regulations to the act and Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 on population and housing censuses (Journal of Laws EU of 13 August 2008, No 218).

The 2011 census covered the persons living permanently (registered for permanent residency) in Poland, regardless of the fact whether these persons were staying in the country during the census or were abroad, as well as persons staying temporarily in the country. The census was taken in buildings, dwellings, collective accommodation places and other housing units.

In Poland, the census practice was so far based on traditional method of taking censuses, which consisted in involving census enumerators visiting all inhabited units and noting down information obtained directly from respondents on census forms, available in a hard copy. Next data from the forms were registered and then in electronic form were subject to further analysis. However, such organization of the census turned out to be very expensive and laborious. For that reason, in census 2011 Poland decided to give up the traditional mode census in favour of a mixed method which consists in combining data from records and information systems with data obtained through direct statistical surveys.

The European Commission, while preparing legal grounds for taking censuses on the territory of the European Union, took into account the dynamic development of administrative systems and IT solutions used when selecting census method. The provisions of Regulation (EC) No 763/2008 of the European Parliament and of the Council provide for specifically defined sources that might be used to collect statistical data for the purposes of the census. Particularly, these sources are as follows:

- a) conventional-censuses,
- b) register-based censuses,
- c) combination of conventional censuses and sample surveys,
- d) combination of register-based censuses, and sample surveys,
- e) combination of register-based censuses and conventional censuses,

- f) combination of register-based censuses, sample surveys and conventional censuses,
- g) Appropriate with rotating samples (“rolling” censuses).

Regulation (EC) No 763/2008 is the first legal act on international scale, which equally treats various approaches to population and housing censuses in the European Community Member States.

## **2. The principle purposes of the census**

The principle purposes of the 2011 Population and Housing Census might be specified as follows:

- 1) fulfilment of the country information needs, particularly obtaining information which cannot be obtained from other sources;
- 2) supply of information at the level of units of basic administrative division of the country;
- 3) the broadest possible description of changes that occurred in the years 2002-2011 in basic demographic and social structures of the population, households and families as well as changes in the size and standards of dwelling stock;
- 4) obtaining information necessary to fulfil international needs of the European Community and the United Nations (UN);
- 5) updating of the base for creating sampling frames for survey samplings conducted through households monitoring.

## **3. The scope of the census**

While establishing the scope of the 2011 Population and Housing Census, the use of previous census results as well as new information needs were analysed. A possibility to obtain information from other sources, mostly from information systems of public administration was taken into account as well as the necessity to keep full comparability of data over time and international obligations provided for in Regulation (EC) No 763/2008.

In 2008, public consultation was held, during which comments and demands for census data were gathered from central and local offices, local authorities, research centres, national and ethnic organizations and religious institutions.

**As a result, the following research topics were established:**

- 1) geographic characteristic of population: place of residence, place of residence in the period between censuses and the reasons for changing thereof;
- 2) demographic characteristic of persons: sex, age, marital status (legal and de facto), country of birth, including the parents country of birth, citizenship;
- 3) households and families: the size and composition of household and family, nuclei and reconstituted families, single-parent families, family status, household status;
- 4) educational characteristic: educational level, the continuation of education, the type of school (educational institution), the field and direction of study;
- 5) internal and international migrations, including emigration of the Poles, labour-related emigration, re-emigration and immigration of foreigners to Poland;
- 6) fertility of women;
- 7) nationality and language;

- 8) religious denomination (a church or religious association);
- 9) legal and biological disability;
- 10) economic characteristics of persons, including:
  - ✓ current activity status – persons having the main and additional job, the unemployed, persons economically inactive,, occupational characteristics of the employed
  - ✓ usual activity status in respect to persons employed on individual agricultural holdings;
  - ✓ commuting to work
- 11) main and additional source of -livelihood of persons
- 12) household sources of economic and living self-dependence

Censuses of dwellings and buildings in which dwelling are located, constitutes an integral part of population census in Polish practice in this respect. Information gathered during the census enable an estimation of housing needs resulting from both actual lack of dwellings and due to the necessity to exchange the existing housing stock or stock not suitable for renovation.

**The following housing-related issues were presented in the census:**

- 1) the type of living quarter;
- 2) the characteristics of dwellings including:
  - ✓ occupancy status of dwellings, type of ownership, size of dwellings including: the number of rooms (1), specifying rooms (2), kitchens and other facilities as well as usable area of dwellings, technical and sanitary system (fittings), method of heating a dwelling;
  - ✓ uninhabited dwellings, additionally described by their purpose and the reasons why such dwellings are uninhabited;
- 3) the characteristics of buildings with inhabited dwelling premises, including: the type of building, type of ownership, the number of dwellings in the building, and period of construction;
- 4) information on tenure status.

#### **4. Sources of data**

The 2011 Population and Housing Census was taken with the use of so-called mixed method i.e. allowing for data acquisition from administrative sources (registers and information systems) as well as their collection directly from population as part of sampling survey and so-called complete survey. Additionally, two complete surveys were conducted, covering persons staying in collective living quarters as well the homeless. The purpose of such solutions was mostly to reduce the census costs and the level of bias with respect to persons covered by the census, without affecting high quality of the census results.

The Act on the 2011 national population and housing census stipulates that the information systems of public administration shall be used as widely as possible as both the source of data for the census purposes (which as a consequence meant that information to be gathered during the census was mostly obtained from available administrative sources and then used to prepare and update an address and housing register followed by preparation of an address and housing frame for samples to be used in sampling survey) as well as direct source

of census data. Data not included in the information systems of public administration or data ineligible in terms of the statistical data quality were collected from persons covered by the census. In this case however, it is envisaged that modern techniques of data collection shall be used in order to eliminate paper forms.<sup>1</sup>

## **5. Full-scale survey**

The survey was conducted with using the Internet – each person could fill out a special electronic form by verifying the data obtained from administrative records and information systems.

As a result of detailed analysis of the collected data, the gathered information was combined, thereby forming a subjective register needed to conduct a census survey. Data acquired through this method were also used to generate prompts in census questionnaires thereby speeding up the process of gathering information from respondents (a short form included 16 questions). About 10% of the census population used this method.

Data on persons who did not participate in the census by the Internet (they did not participate in the sample survey and in the survey in the collective living quarters or the survey of the homeless) were generated on the basis of the information available in the administrative sources (registers and information systems).

Using administrative data was related to the need to implement procedures for standardization, synchronization and validation. First of all registers and information systems that contain reference data were extracted

The so-called building resource was the main source of information on buildings and dwellings in the 2011 census. This resource was created as a result of combining data collected within the 2002 National Population and Housing Census, statistical surveys regarding, among others, building permits issued, residential buildings and dwellings in completed inhabited buildings as well as the Building Electronic File. This File was used by each entity managing or administrating buildings to pass on information regarding selected features and parameters of particular multi-dwelling buildings.

## **6. Sampling survey**

Sampling survey conducted as part of the 2011 Census provided data which are not collected in registers and information systems. The survey was conducted on random sample of approx. 20% of dwellings on national scale. A dwelling, precisely its address, was a sampling unit. A collection of dwellings, that was the basis for the sampling was prepared as the sampling frame with the "deep" stratification.

The sample frame required a special preparation due to the fact that it was applied the one-step stratified sampling scheme, the corresponding sampling frame and the sample allocation in various poviats (in all previous censuses in the accompanying census sample surveys used two-steps stratified sampling scheme.<sup>2</sup>

---

<sup>1</sup> More information on the sources of the data used for the 2011 Census are presented in Chapter I in the "Report of the results. National Census of Population and Housing 2011 ", GUS, Warsaw 2012.

<sup>2</sup> Preparation of the sampling frame, sampling scheme and sample allocation were examined in detail in Chapter 1 of the publication "Methodology of Population and Housing Census 2011 – Selected Aspects". Moreover, a

**In the end, these actions resulted in sampling over 2744 thous. dwellings out of almost 13.5 million dwellings. Almost 70.5 thousand strata were created whereas the sample size in given strata fluctuated from almost 6% to over 49%.**

The scope of sampling survey in the 2011 Census covered six big thematic areas:

- population and its demographic and social characteristics,
- economic activity,
- internal and international migrations of population,
- nationality and religious denomination
- households and families
- buildings and dwellings.

15 research topics might be distinguished within these areas. A long form, of broad range of topics with many questions (over 120 questions) was used in sampling survey.

Respondents answered to about 70-80 questions, depending on his/her sex, age, mobility and economic activity. Sample survey in the vast majority were carried out by direct interviews with the residents of dwelling (CAPI method), but respondents could also decided on self-enumeration by the Internet (CAII method) – about 2% persons have used this method.

## **7. Census of persons in collective living quarters and the homeless persons**

During the census, information on persons staying in collective living quarters for over 3 months, i.e. places occupied by one separate institution providing services for chronic medical care, nursing and health, as well as buildings related to work or education (student dormitories, boarding houses, workers' hostels) or others in which a large number of persons live/stay on the usual basis was gathered. Data were collected from the building's owners, administrators and stewards with the use of special Internet application. In special cases the information was gathered with the help of the voivodship and gmina census offices personnel and the census enumerators.

**In 15-16 April 2011, in cooperation with the Pomeranian Forum in Aid of Getting Out of Homelessness, the survey of the homeless was also conducted.**

The homeless persons were recorded by a census enumerator with the use of mobile application wherever the homeless were staying, as indicated by the gmina census offices in agreement with institutions proving aid to the homeless. Persons spending the census evening and night outside any of round-the-clock institutions in such places as: railway and bus stations and their neighbourhood, heating ducts and units of district heating system, allotments, streets, bunkers, woods, parks, shopping centres, car parks, abandoned cars, caravans, staircases, rubbish chutes, cellars, rubbish tips, carriages, sidings, places where the homeless might warm up etc. were subject to the census and were recorded for its the purposes. The homeless staying in collective living quarters (hostels, night shelters and institutions for the homeless) were recorded by these facilities administrators.

## **8. The pilot census**

The pilot census was taken in the period from 1 April to 31 May 2010 as of 31 March 2010, at 00.00.

Units of territorial division, in which the pilot census took place, were selected intentionally, taking into account their diversity in terms of population density, distinctive features of lands, population that lives there and its housing conditions. On the territory of the indicated units of territorial division, all inhabited buildings and dwellings as well as the persons were recorded according to the rules accepted with respect to the Census 2011. Methodological and organizational solutions to be used in the Census 2011 were verified. Individual methods of data collection were verified with particular precision, including the efficiency of self-enumeration by the Internet and telephone interviewing. Testing an application to be used with respect to electronic form and the efficiency of data acquisition on portable electronic devices as well as technical issues related to uploading completed electronic forms to the right servers were equally important.

## **9. The post enumeration census**

From 1 to 11 July 2011 the post enumeration census was taken for the 2011 Polish Population and Housing Census. The purpose of the 2011 post enumeration census was to check the completeness of the census taken, the correctness of data collected during the census as well as conformity of data to the factual state.

Out of 2.744 thousand dwellings that had been previously sampled for sampling survey, 80.000 thousand dwellings were sampled. The following dwellings were subject to the test census: dwellings where residents decided on self-enumeration by the Internet, dwellings which were recorded directly by census enumerators as well as dwellings recorded on the phone by interviewers and these for which the census was not taken (for various reasons). The post enumeration census was taken by interviewers on the phone (the CATI method). The post enumeration census form included 14 questions.

## **10. Forms used for the purposes of the 2011 census**

Two types of forms available only in electronic form were used for the purposes of the 2011 census. A long form, of broad range of topics with many questions (over 120 questions) was used in sampling survey, whereas a short form (16 questions) was used in full-scale survey, mostly in order to update data from registers and information systems. Electronic forms were available on-line, and a short form was also available off-line. The forms were prepared for use by an application designed for portable devices of the hand-held type as well as for Internet application which was used during self-enumeration by the Internet. The electronic application was provided with glossaries for questions regarding particular thematic areas such as: education, economic activity, the country of citizenship and birth, migration, national or ethnic identity, religious denomination. Some glossaries allowed free verbal record. Additionally, in the address part of electronic forms, the TERYT glossary was attached.

Separate, simplified electronic forms were drawn up with regard to the survey of the homeless and the persons staying in the collective living quarters.

As part of the 2011 census, the complete survey of population in 86 gminas initially selected on the basis of the 2002 census results was conducted. The main criterion allowing gminas to be selected was at least 10% share of persons belonging to national or ethnic minority per given gmina population in 2002. With regard to persons living or staying in such gminas in sampled dwellings, a long form was filled in, whereas in other dwellings – short form was applied. The questions regarding national or ethnic identity as well as the language used at home were added to both forms since none of administration systems included information on nationality, which could be used in the population census. Data from this survey are of essential importance for drawing up the census results in the area of nationality and language, particularly in order to determine so-called minority gminas.

## **11. Methods and forms of dissemination of the census results**

The following methods shall apply for dissemination of the 2011 population and housing census results:

- tabular and analytical publications as well as analytical publications,
- dissemination of data on data carriers,
- dissemination of data through the Internet,
- dissemination of data through the Local Data Bank (BDL)
- dissemination of data through thematic (field) databases e.g. the DEMOGRAFIA database,
- dissemination of data through direct access to final statistical information – in Analytical Microdata Base (ABM)

Most of the operations connected with the analysis and dissemination of the census results take place within the ABM system. Internal users shall have full access thereto, also to individual data (non- identifiable) of the 2011 census, whereas external users' access shall be monitored by special application.

One might separate the main processes supported by the ABM system, i.e. data processing, data analysis and data dissemination. It is assumed that as part of data dissemination process, the following actions will take place in the ABM system:

- ✓ preparation of products for dissemination,
- ✓ management of disseminated products,
- ✓ monitoring and analysis of the users' inquiries.

The SAS system shall be used to support the ABM system, however it will be possible to combine results prepared outside the ABM e.g. in the SPSS application or in Excel spreadsheet. A module supporting the process of monitoring and analysing the users' information needs with regard to census data shall play an important part in data dissemination. Feedback received thanks to this module might be used to modify the trends in dissemination of the 2011 population and housing census results.

Depending on the method of dissemination of the census results, various forms of their presentation will be used:

- ✓ data aggregated in public predefined tables
- ✓ data created by the users on the basis of micro aggregates
- ✓ data developed by statisticians upon request

- ✓ results of analytical research conducted on data included in the OLAP cube or on the collection of individual non-identifiable data
- ✓ graphic presentation of data (graphs, cartograms)
- ✓ visualisation of the census results with the use of the GIS tools through the GSP (geostatistical portal).

Various forms of the census data dissemination, particularly extensive set of published or predefined tables, available in the ABM system as well as in other bases, should fulfil the basic needs of a wide circle of users of the census results at the national and regional level. Predefined objects, i.e. aggregates and multidimensional cubes, shall constitute the basic source of data for both external and internal users. It is assumed that calculating approx. 90% of statistical tables should be possible based on previously prepared aggregates in the OLAP cubes.

Those users of the census results who are literate in IT solutions will have a opportunity to calculate simple correlation tables independently, through the access to the ABM as well as to the Meta Data Subsystem (PM). Access to databases shall be governed by the principle of personal data protection; the presented data cannot be identifiable. However, tables requiring data to be processed (calculated) upon recipients' individual, special requests including atypical territorial sections, wider data correlation or unusual groupings shall be prepared by specialized statistical units.

## **12. Publishing of the 2011 census results**

Tabular and analytical publications to be published might be presented in the following two groups:

- ✓ national (including data for the country and for regions, specifying urban and rural areas, and voivodships);
- ✓ regional (including data for voivodships, sub regions, powiats, specifying urban and rural areas, and also important information for gminas in each of voivodships);

In 2012-2013 the following thematic publications will be issued:

1. Population. Size and socio-demographic structure
2. International migration
3. Occupied buildings
4. Dwellings
5. Economic activity of the population
6. Households and families. Demographic characteristics
7. Population and family households – socio-economic structure
8. Internal migration
9. Selected aspects of economic activity of the population
10. Composition by national and ethnic identity, language and religion denomination of population
11. Housing arrangements of households and families

In addition, the regional publications (reports and other publications) will be developed and published by all statistical offices. These studies will focus on issues of population, migration, economic activity and housing in each of the voivodships.

## **II. THE METHODOLOGY OF THE 2011 NATIONAL POPULATION AND HOUSING CENSUS – selected aspects**

### **1.1. Census definitions and terms**

This part of the publication features the definitions which are necessary to properly interpret the Census results as regards the population and its socio-demographic characteristics.

#### **Population categories**

Based on the results of the 2011 National Population and Housing Census, the following population categories have been distinguished:

1. De jure resident population (actually residing population)
2. Residents (residing population).

Information presented in all tables contained in the publication concerns the category of actually residing population, hereinafter referred to as “the population”.

**Actually residing population (actual population).** This category, from the point of view of the residents of a gmina, i.e. the smallest unit of administrative division of the country, comprises the following groups:

1. **Actually residing persons** – these are persons who have declared in the Census that a given gmina is their permanent place of residence (irrespective of whether they have been registered for permanent residence there or not), and who at the moment of the Census date were:
  - a) present (residing in the gmina),
  - b) absent, but their absence was shorter than 3 months,
  - c) were absent for more than 3 months, and their absence was due to the following reasons:
    - staying in the penal institute or in custody, staying abroad.
2. **Persons temporarily staying for more than 3 months.** This concerns persons who declared in the Census that their permanent place of residence is in another gmina in the country whilst at the place of enumeration they are residing temporarily for the following reasons: education, work, family or housing conditions, rehabilitation or treatment,... dwelling in the welfare establishment

The period of temporary absence or stay was assumed to be an intended period.

The category of actually residing population of a given gmina excludes persons coming from abroad for temporary stay, i.e. persons not registered for permanent residence in Poland (without a settlement permit). This category also excludes permanent residents of a given gmina who at the Census date stayed away for more than 3 months for other reasons than staying in the penal institute (or in custody) or staying abroad, i.e. persons temporarily staying

in other gmina in the country (in accordance with the definition adopted, they were included there as actually residing persons). **Actually residing** population of individual localities is determined in the same way.

**Residing population (residents).** From the point of view of gmina residents, residing population comprises:

- a) usual residents, excluding persons staying away from their place (gmina) of residence for at least 12 months, regardless whether their place of stay is in the country or abroad. Polish diplomats staying in diplomatic posts and soldiers on foreign military missions constitute an exception as they were enumerated as permanent residents of Poland (in accordance with the UNECE and EU recommendations) in spite of staying abroad,
- b) Persons temporarily staying in the gmina for at least 12 months, coming from another locality in the country or from abroad (foreigners who are not registered for permanent residence in Poland).

Criteria of population migration when marking out categories of residents are as follows: education, work, family and housing conditions, medical treatment and rehabilitation, dwelling in the welfare establishment. This means that persons being an inmate of the penal institute or in custody – irrespective of the duration of absence – are considered residents of the locality in which they resided before they were “forced” to leave.

### **Age**

The age of a persons – expressed in the number of years completed – was determined by comparing full date of their birth with the date of enumeration (referred to as the critical date, i.e. 31 March 2011).

### **Marital status**

Considering the legal regulations binding in various countries, regarding the minimum age for marriage, it was decided that marital status in population censuses would be determined for persons aged 15 years or more.

#### **Legal marital status**

The legal marital status of a person is defined as the legal status of a given person in relation to marriage or civil union in accordance with laws or customs of the country (de jure status). In accordance with the Polish law, there are four categories of marital status:

- single – regarding persons who have never been legally married,
- married – regarding persons who have entered into a civil-law marriage,
- widowed – regarding persons whose legal marriage ceased to exist because of the spouse’s death,
- divorced – regarding persons whose marriage was dissolved by court.

Persons for whom the court has adjudicated separation are still considered married, from the point of view of the law.

#### **The facto marital status**

The facto marital status was determined secondarily, on the basis of legal marital status, the nature of given person's the relationship, i.e. on the basis of information pertaining to the

relation with the reference person of household and mutual relations between the enumerated persons

The following categories of the actual marital status have been distinguished:

- single, i.e. persons who have never been married and who at the census date did not live in cohabitation with any other person,
- married – persons who have entered into legal marriage and the marriage still actually exists. Such persons declared that they formed a community by marriage, irrespective of whether the spouses were enumerated together (in the same dwelling) or separately (e.g. in the case of the absence of one of the spouses due to education, work or the lack of a joint dwelling).

Persons who remained to be legally married but who did not form a community by marriage, under a decision made by one or both spouses, and whose marriage was not dissolved by court (by the adjudication of divorce or separation), were not considered married, and their de facto marital status was determined, in accordance with the respondent's declaration, as separated or living in cohabitation with another person,

- partner – cohabitation was determined to exist within the same household, irrespective of the legal marital status of the persons living in cohabitation,
- widowed – persons whose legal marriage ceased to exist because of the spouse's death, and who at the moment of Census did not live in cohabitation with another person,
- divorced – persons whose marriage was dissolved by court, and who at the moment of Census did not live in cohabitation with another person,
- separated – this category refers to persons who at the moment of Census:
  - a) were legally separated and did not live in cohabitation with another person,
  - b) were still legally married but neither formed a community by marriage nor lived in cohabitation with another person.

### **Country of birth**

In accordance with international recommendations, the country of birth had to be specified by referring to the state borders valid at the moment of Census, and not on the day of birth, e.g. if a person was born in Vilnius, the country of birth should be Lithuania, regardless of the year of birth. Such a solution was applied with a view to ensuring comparability, and especially to avoiding data duplication on the international scale.

### **Citizenship**

Citizenship is defined as a particular connection (legal bond) between a person and the State. It does not show the ethnic origin and is not depended on nationality.. A person may have one, two or more additional citizenships. It is also possible that a person has no citizenship. A

person having the citizenship of Poland and of any other country in Poland is always treated as a Polish citizen, and never as a foreigner.

**A foreigner** is any person who does not have Polish citizenship, irrespective of having (or not having) a citizenship(s) of any other country (countries). Foreigners coming to Poland for permanent residence are considered permanently residing persons, whereas others, depending on the actual duration of stay in our country, are treated as:

- temporarily staying (short-term immigrants) – in the case of stay for less than a year,
- residents (long-term immigrants) – in the case of stay for at least 12 months.

**A stateless person** is a person with no citizenship (stateless persons are considered foreigners).

### **Nationality or ethnic identity**

– is a declared (based on a subjective impression) individual feature of each person expressing his/her emotional or cultural relationship, or the one following from his/her parents' origin, to a specific nation or ethnic community.

For the first time in the history of Polish censuses, the 2011 census allowed the inhabitants of Poland to express complex national and ethnic identities by asking the respondents two questions regarding national and ethnic affiliation.

**Language spoken at home** - the language used in everyday contacts with the closest persons. In census 2011 a solution has been adopted which enables respondents to register two names of non-Polish languages regardless of the fact, whether they were spoken jointly with Polish language, or exclusively.

### **Mother tongue**

The mother tongue should be considered the language which a person learnt as first in his/her early childhood (which he/she learnt to speak), or possibly the language which his/her parents or custodians used the most often when speaking to him/her in his/her early childhood.

### **Religion – religious affiliation**

Religion (religious affiliation) refers to formal participation in, or emotional relation of a person with a given religion (church or religious denomination). Religion should be determined on the basis of a voluntary declaration, including the declaration of the lack of relation with any religion (no religious affiliation).

### **Education**

This is the highest educational level achieved in a school-based system, or in any other training mode or form, recognised in compliance with the binding education system. The basis for classifying education to a given level is provided by a certificate (diploma) of completion of a certain school, irrespective of the mode of studying (full-time, evening/weekend, part-time). Information on the educational level was obtained for all persons aged 13 years or more.

In the 2011 Census , as compared with the 2002 Census , the classification of educational levels was extended by two items, i.e. graduates holding a college diploma (at the post-secondary school level) and lower-secondary school leavers.

### **The classification of educational levels**

#### **Higher education:**

1 – refers to persons with a degree of Doctor (PhD), Assistant Professor (PhD hab.) or Professor,

2 – refers to persons with a degree of Master of Arts/Sciences (MA, MSc), Medical Doctor or equivalent, obtained upon completing second-degree studies (supplementary Master studies) or uniform Master studies,

3 – refers to persons with a degree of Engineer, Bachelor or certified economist, obtained upon completing first-degree studies (Bachelor or Engineer studies).

#### **College diploma:**

4 – refers to graduates from teacher training colleges and teacher training colleges of foreign languages (except when organised as part of a higher education establishment), and to graduates of colleges of social work. This category also includes persons completing post-secondary vocational schools for teachers.

#### **Post-secondary education:**

5 – post-secondary with “Matura” certificate – refers to persons holding a diploma (certificate) of completing a post-secondary school, the admission to which was conditioned on holding a “Matura” certificate,

6 – post-secondary without “Matura” certificate – refers to persons holding a diploma (certificate) of completing a post-secondary school, the admission to which was conditioned on holding a certificate of completing a secondary school.

#### **Secondary education:**

7 – secondary vocational with “Matura” certificate – refers to persons who obtained a “Matura” certificate in a secondary vocational school (a secondary technical school, a supplementary technical school, a secondary vocational school, or a second-degree art school),

8 – secondary vocational without “Matura” certificate – refers to persons who obtained a certificate of completing a secondary vocational school (a secondary technical school, a supplementary technical school, a secondary vocational school, or a second-degree art school),

9 – general secondary with “Matura” certificate – refers to persons who obtained a “Matura” certificate in a general or specialised secondary school, or who completed a lower secondary school before 1932 (*persons who completed a lower secondary school in 1932-1948 (a 4-year lower secondary school) were classified at a primary level*),

10 – general secondary without “Matura” certificate – refers to persons who obtained a certificate of completing a general or specialised secondary school.

**Basic vocational education:**

11 – refers to persons holding a certificate of completing a vocational school (a basic vocational school, a farming school, an industry school, a vocational lower-secondary school, etc.), a preparatory vocational or farming school, or a correspondence course in farming, exclusively at the level equivalent to a basic vocational school or a certified master school.

**Lower-secondary education:**

12 – refers to persons who obtained a certificate of completing a lower-secondary school (lower-secondary schools became operative in the 1999/2000 school year). This category also covers graduates of special job-training schools.

**Primary education (complete):**

13 – refers to persons holding a certificate of completing a primary school (before the war a mainstream school), irrespective of the number of classes (six, eight or four as in the past), courses for working persons at the level of primary school or a first-degree art school which subsequently covers the primary-school curriculum.**Incomplete primary education and without school education:**

14 – refers to persons who currently attend, or used to attend but did not complete, primary school, or who have never attended primary school.

**Continuation of education**

This refers to attending schools in various modes, i.e. full-time, evening/weekend or part-time, the completion of which will increase the current educational level of a person.

New issues added to the education module in the 2011 Census include **the field of education – regarding persons holding a scientific title or degree, specialty – regarding persons holding a diploma of completing Master or Bachelor studies, and trained occupation – specified by graduates from post-secondary, secondary, vocational or basic vocational schools**. Information was also collected on the type of school in which a respondent continues his/her education, based on the following classification: PhD studies, Master studies, Bachelor studies, colleges and post-secondary schools, technical secondary schools, general and specialised secondary schools, basic vocational schools, lower-secondary schools and primary schools.

The following dictionaries were used in the National Population and Housing Census 2011, in order to determine the field of education, specialty and trained occupation: „*Klasyfikacja dziedzin nauki i dyscyplin naukowych*” (The classification of the fields and disciplines of science), „*Klasyfikacja kierunków studiów*” (The classification of the fields of studies) and „*Klasyfikacja zawodów*” (The classification of occupations). These dictionaries were prepared in accordance with ISCED 97 (*the International Standard Classification of Education*).

In the 2011 Census, the results in this area were prepared by referring to the following groups: 100, 200, 222, 300, 420, 440, 460, 481, 500, 600, 700, 800, 900.

The following areas were considered as part of the groups listed:

- 100 – *Teacher training and education science* – this group comprises general and specialised teacher training programmes, and also education science;
- 200 – *Humanities and arts* – this group comprises fine arts, performing arts, design and (artistic) craft skills, religion, Polish philology, history and archaeology, philosophy and ethics;
- 222 – *Foreign languages* – linguistics;
- 300 – *Social sciences, business and law* – this group includes, among others, psychology, sociology, political sciences, economics, journalism and social communication, marketing and advertising, finance, banking and insurance, bookkeeping and taxes, trade and law;
- 420 – *Life sciences* – this includes biology, biophysics, biochemistry, microbiology, medical biology and ecology;
- 440 – *Physical sciences* – this includes physics, chemistry, geography, geology, geophysics, astronomy and marine science;
- 460 – *Mathematics and statistics*;
- 481 – *Computing*;
- 500 – *Engineering, manufacturing and construction* – this group comprises, among others, the following specialties: machine and metallurgy industry, electricity and energy, electronics and automation, chemical processes, mechanical vehicles, production and processing, mining and quarrying, construction;
- 600 – *Agriculture and veterinary* – this category includes crop and livestock production, horticulture, forestry, fisheries and veterinary medicine;
- 700 – *Health and welfare* – this group comprises medicine, nursing, medical diagnostics, therapeutics and rehabilitation, pharmacy, child care, youth services and social work;
- 800 – *Services* – these are, among others, personal services for the population, hotel and catering, travel and tourism, leisure, sports and physical culture sciences, environmental protection, protection of property and persons, civil security, health and safety at work, military forces and national defence;
- 900 – *Not known or unspecified* – classification to the right category is impossible.

As regards the field of education, a group of foreign languages (222) was not separated, whereas for trained occupation the classification by group does not cover foreign languages (222), life sciences (420), and mathematics and statistics (460).

## **Dwelling**

A dwelling is premises comprising one or several rooms, including auxiliary rooms, built or remodelled for residential purposes, structurally separated (by fixed walls) within a building into which a separate entrance leads from a staircase, a passage, a common hall or directly from the street, courtyard or garden.

## Homeless person

A homeless person is a person who declares, for various reasons, e.g., economic, family-related or administrative, that he/she does not have any permanent place of residence. This category also comprises persons who have been deprived of their dwelling due to fortuitous events (cataclysms, floods, fires, etc.).

**Collective living quarters** This is a complex of rooms and other auxiliary facilities located in one or more buildings, and occupied by one separate establishment rendering the following type of services: care and education, residential healthcare, hotel services or other, and which the services is usually inhabited by a larger number of persons. If in one building two or more separate entities were located, e.g. a children's home and a congregation of nuns leading the children's home, a hospital and an employee hotel for nurses, a boarding house for theological seminary students and a congregation of priests, each entity was enumerated on a separate basis. In residential and boarding facilities which features a registered boarding house (a hotel, a holiday house, etc.), those rooms which were intended for permanent residence of the building owner, administrator or employees running separate households were considered a dwelling. The rooms intended as a boarding house (a hotel, or a holiday house) were considered collective living quarters

### The following classification of collective living quarters was applied

| Symbol | Types of collective living quarters                                                                                                                                                                                                                |
|--------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 51     | Boarding schools and dormitories                                                                                                                                                                                                                   |
| 52     | Student dormitories, boarding houses for junior members of teaching staff (unless converted into dwellings) boarding houses for theological seminary students                                                                                      |
| 53     | Employee hotels                                                                                                                                                                                                                                    |
| 54     | Children's homes, family-based children's homes, children's villages                                                                                                                                                                               |
| 55     | Other care and education facilities ( <i>except for juvenile detention centres and juvenile shelters</i> ) including children's shelters, special purpose school and education centres, youth educational centres, child support facilities        |
| 56     | Social assistance homes for retirees and elderly persons                                                                                                                                                                                           |
| 57     | Social assistance homes for single women in pregnancy or with little children                                                                                                                                                                      |
| 58     | Facilities for chronically ill persons suffering from somatic diseases, for chronically ill persons suffering from mental diseases, for mentally-handicapped persons, for persons with physical disabilities, youth sociotherapy centres, hospices |
| 59     | Mental facilities (hospitals), residential treatment centres, rehabilitation centres for drug addicts                                                                                                                                              |
| 60     | Convents, monasteries, clerical religious congregations (including habitless)                                                                                                                                                                      |
| 61     | Homeless shelters, Homeless hostels                                                                                                                                                                                                                |
| 62     | Other collective living quarters, including hotels, motels, boarding houses, holiday houses, tourist shelters, campsites, hospitals, sanatoriums, preventoriums.                                                                                   |

### The classifications and groupings used for population in NSP 2011

1. The classifications and groupings defined in Commission Regulation (EC) No. 1201/2009 of 30 November 2009 implementing Regulation (EC) No. 763/2008 of the European Parliament and of the Council on population and housing censuses in terms of technical specifications of the topics as well as of their breakdowns
2. The International Standard Classification of Education (ISCED – UOE, 1997)
3. The International Standard ISO 3166-1, Codes for the representation of names of countries and their subdivisions-Part 1: Country codes, ISO 3166-1: 2006 (E/F), International Organization on Standardization (Geneva, 2006)

## **1.2. The quality of the full census and sample survey results**

The activities related to establishing an electronic list of addresses and dwellings for Census purposes constituted an important area of the preparatory work performed as part of 2011 Census. This stage of work was crucial to the census conducted with modern methods. The register of addresses and dwellings was the source of data for *the Master Record (MR)*, to be used at all Census stages, for the purpose of determining the Census completeness, also serving as the basis for establishing the sampling frame to be used in the sample survey.

The list of addresses and dwellings was prepared on the basis of the National Official Register of Territorial Division of the Country (the TERYT register), along with data coming from other sources, such as the national geodesic and cartographic resources, as regards the spatial location of buildings. As a result of combining these two sources, while establishing the address identifiers of buildings, the geographic coordinates x, y (address points) from the TERYT register were also considered. As regards the information concerning population, the register of addresses and dwellings made a reference to , gmina registration collections, which were combined with the NOBC system, i.e. the address identification system of streets, real property, buildings and dwellings, forming part of the TERYT register. The resultant compilation of buildings, dwellings and persons was then verified by gmina offices as part of pre-census updates. This stage entailed determining the population of buildings, dwellings and collective living quarters to be enumerated, along with verifying the accuracy of addresses, and assigning persons to specific dwellings and places.

In the course of the pre-census survey, the census enumerators verified all address points assigned to specific census areas, comparing them with the items included in the list of addresses and dwellings used in the pre-census survey. The aim of this interview was to either confirm or modify the address points included in the list, and to remove the non-existent points or to possibly add the new ones which were not included in the list, which comprised residential buildings, non-residential buildings with dwellings or collective living buildings, e.g. student dormitories. In contrast to previous censuses of population and housing, the enumerators did not visit dwellings in the pre-census survey. It was therefore impossible for them to verify the information concerning the main occupier of a dwelling and the number of persons to be enumerated.

### **The dwelling sampling frame used in the sample survey**

The establishing of the sampling frame comprised the following stages:

1. Establishing an integrated register of addresses and dwellings, based on the Address and Housing Specification;
2. Verifying the register in the course of the pre-census survey;
3. Supplementing the register with additional variables coming from other registers and information systems, along with establishing the sampling frame to be used in the sample survey.

Information coming from other registers and information systems, or statistical databases, was also included in the register of addresses and dwellings (as part of the Operational Micro Database). These sources included the datasets of the Social Insurance Institution (the Central

Register of Insured Persons, the Central Register of Retirees and Pensioners, and the Central Register of Contribution Payers), datasets of the Agricultural Social Insurance Fund (KRUS), the National Taxpayers Record, the PESEL register, the 2002 National Population and Housing Census, building resource and the Base Register for the 2010 Agriculture Census. The aim of including additional information in the list was to improve the quality of data gathered and to ensure all information necessary to stratify the sampling frame. In the course of data verifications, certain criteria (variables and their values) were adopted, which were then used to establish more homogenous groups (strata) of survey units.

The following items were excluded while establishing the sampling frame:

- a) collective living quarters not including any dwellings,
- b) locked facilities and dwellings located therein,
- c) Dwellings in collective living quarters constituting locked facilities,
- d) semi-permanent housing unit,
- e) uninhabited dwellings which were damaged during natural disasters, including especially floods.

As a result, the sampling frame covered dwellings occupied by persons registered or non-registered for permanent or temporary stay, together with unoccupied dwellings which were being repaired, in which tenants were changing, or newly-constructed dwellings. Buildings recorded under the same address were considered in the sampling process unless their number exceeded two. The dwelling sampling frame was updated in the course of the pre-census survey, as a result of which some buildings which were not previously included in the register were also added by the survey.

### **The variables included in the sampling frame**

For the purpose of the sample survey, the reference list was supplemented with a set of information concerning each dwelling, which was then transferred to the sampling frame, constituting the basis for the ultimate selection of the stratification criteria:

- a) The territorial symbol,
- b) The building/dwelling location in the area affected by a natural disaster,
- c) The number of dwellings in a multi-dwelling building,
- d) The location of a dwelling in a multi-dwelling building,
- e) The year of completion of building construction,
- f) The number of persons in a dwelling (0, 1, 2, 3, 4, 5, 6, 7 or more) ,
- g) The number of persons registered for permanent residence,
- h) The presence of a foreigner in a dwelling,
- i) The presence of a working (insured) person, a retiree or a pensioner in a dwelling, or the presence of an unemployed person,
- j) The presence of a disabled person,
- k) The presence of an agricultural holding user and a utilized agricultural area,

The location of dwellings in a gmina, with the participation of persons belonging to national-ethnic minorities share amounting to at least 10%.

The values of variables, as the stratification criteria, were diversified for urban and rural areas. The choice of criteria was finally determined in the course of trial sampling.

### **Non-sampling errors in the 2011 National Population and Housing Census**

The results of the 2011 Census verification constitute the principal source of data for the purpose of assessing the occurrence and scale of non-sampling errors. Data gathered in the census verification process should especially allow assessing coverage errors (both under- and overcoverage) resulting from the failure to enumerate some dwellings and persons, which might have not been included in the sampling frame, from the omission of some persons who should have been enumerated, from a double enumeration of persons, or from a possible addition of fictitious persons. Such data could also be used to determine the content errors resulting from non-response or from response errors. The assumptions made and the data gathered in the census verification process should further make it possible to assess the impact of external factors, such as enumerators, interviewers or respondents who were self-enumerated, on the quality of results of the National Population and Housing Census 2011.

Non-sampling errors may also stem from data obtained from registers and information systems, due to the lack of integration between various databases. Therefore, such data required special validation which comprised normalisation, conversion, as well as subject, object and spatial synchronisation procedures. It is of utmost importance to thoroughly examine the definitions and terms applied in various registers and information systems, which arise from the provisions of specific legal regulations that provide the basis for such registers and systems.

The effects of an in-depth analysis of potential non-sampling errors, occurring both in the full census and sample surveys, will be presented, where possible, in future national publications. Nevertheless, it should be borne in mind that such errors are extremely hard to detect, whereas an estimation of their impact on the census result is even more difficult, if not impossible. Detailed methodological principles of census verification, along with the results, will be presented in a publication devoted to the methodology of the National Population and Housing Census 2011.

### **Sampling errors**

Sampling errors occur exclusively in sample surveys and are always connected with the sampling scheme. They are easy to estimate whenever there is a full compliance with representative method principles, including especially the well-prepared sampling frame and the proper selection of sampling units (cf. Subsection 1.3). The sampling error for the sample survey conducted as part of the 2011 Census was determined on the basis of estimation precision rates (cf. Subsection 1.6 *ibid.*).

### 1.3. The sampling scheme and units allocation in powiats<sup>3</sup>

The principal aim of the sample survey conducted as part of the 2011 Census was to obtain information on the socio-demographic situation at the powiat level. It was assumed that in order to obtain representative results at the level of approx. 400 units in the territorial profile (including 380 powiats), a dwelling sample of approx. 20% in national terms should be selected. A one-level stratified sampling scheme was used with a view to selecting the sample with a given number of units. Before the sampling process began, the sampling units (dwellings) were grouped into strata in order to increase the sampling efficiency. A differentiated approach to stratification was applied, depending on the powiat and gmina type.

**In urban areas – powiats, and in districts and representations of the largest five cities,** and also in other specified larger cities without powiat status, the first stage entailed dividing all dwellings into two categories:

- a) Dwellings in blocks of flats,
  - b) Other dwellings.
- 
- a) An indicator of the number of dwellings in the building in which a given dwelling was located served as the criterion for dividing dwellings into the reference categories. The category of dwellings in blocks of flats comprised all dwellings for which the value of the reference indicator exceeded the median. Then, in each of these two categories dwellings were stratified according to the number of persons in a dwelling, which was followed by a further division into four groups in terms of: The presence of a working person in a dwelling,
  - b) The presence of a retiree or pensioner in a dwelling, in case of absence of a working person,
  - c) The presence of an unemployed person in a dwelling, in case of absence of persons mentioned in (b),
  - d) Dwellings with other persons.

Dwellings located outside blocks of flats additionally comprised the category of a dwelling with an agriculture holding user

Stratification according to the number of persons residing in a given dwelling was a crucial factor contributing to the precision of the results, as it mitigated the negative effects of a diversified number of dwelling occupants. The efficiency of this stratification stage depended on the correlation between the number of dwellings occupied, as entered in the sampling frame, and the actual status. In turn, stratification in terms of other variables had a positive impact, among others, on the census results related to economic activity. **In other powiats, the first stage comprised stratification by gminas, where urban and rural gminas were considered two separate gminas.** In towns, the stage of dividing dwellings into dwellings in blocks of flats and other dwellings was excluded, and the stratification proceeded in the same way as in the case of dwellings located in big cities outside blocks of flats in cities.

**In rural gminas,** the first stage entailed dividing dwellings into two categories:

---

<sup>3</sup> A working paper prepared by a group of mathematicians under the supervision of Bronisław Lednicki.

- a) Dwellings with an agricultural holding user,
- b) Other dwellings.

In the first category the stratification was made according to the number of dwelling occupants, followed, where possible, by further stratification by an agricultural holding area (two or three area groups depending on the number of dwellings with an agricultural holding user in a gmina). In the second category dwellings were stratified in the same way as dwellings in towns.

In some powiats the stratification concept presented above could lead to establishing very scattered strata. As a result, in the course of establishing the strata, the sampling programme simultaneously checked whether the strata established complied with the requirement of the minimum number of dwellings. Strata comprising insufficient number of dwellings were combined with the adjacent strata. The assumed hierarchy of criteria of so-called strata splicing was reverse to the hierarchy assumed for strata creation

Once all strata were established, the number of samples to be selected in each stratum was determined. It was assumed that the sample will be allocated proportionally to all strata. This implied, among others, that the sampling fraction determined through square root allocation for a given powiat was applicable to all gminas in this powiat, and to all strata established within a gmina.

### **Sample allocation in powiats**

A 20% dwelling sample for Poland was divided into powiats using the root square allocation method which constitutes a compromise method between proportionate allocation and allocation ensuring equal precision for all subpopulations. In the case of using the proportional sampling, the sample in each powiat would account for 20% of the population. As the precision of results, i.e. the size of the sampling error, depends on the number of units in the sample, proportional sampling would result in an insufficient precision for a number of smaller powiats. In turn, the alternative method would result in obtaining roughly the same precision for all powiats at the cost of “condensing” the number of units in the sample. In consequence, the sample size, and thus the labour-intensity of census implementation, would differ too little between large and small powiats. For these reasons square root allocation method was applied as a sample division method: the number of dwellings subject to sampling in particular powiat is proportional to a square root of population number of dwellings and is given by the following formula:

$$n_p = n^* \frac{\sqrt{N_p}}{\sum_p \sqrt{N_p}},$$

where:

$n^*$  – the assumed size of the sample for Poland,

$N_p$  – the number of dwellings in  $p^{\text{th}}$  powiat,

This formula was used to calculate sample sizes for various powiats. Treating Warsaw as any other powiat could cause the inability to obtain fairly accurate results for various districts. Therefore, an average value of 87.5 thous. dwellings was assumed as the final sample for the city of Warsaw. This value was subtracted from the assumed size of the sample for Poland, i.e. from 2 631 thous., and then added to the root square allocation for 378 powiats. Samples in the smallest powiats, i.e. in beskidzki and sejneński powiats, amounted to approx. 3200 dwellings (49% of the population number of dwellings), whereas the largest sample (outside Warsaw) was recorded for the city of Łódź – 23 thous. dwellings, i.e. 6.9% of the number of dwellings in this city with powiat status.

The sample in Warsaw, Łódź, Kraków, Wrocław and Poznań was also divided into various districts and branch offices using the root square method. . Outside Warsaw, there was no need to increase the sample size (just as in the case of Warsaw) in other cities due to small number of quarters (4 or 5). The sample sizes determined for particular quarters in these cities ensures the results to be at least as precise as in the smallest powiats.

Taking those measures resulted in selecting a sample comprising over 2744 thous. dwellings out of nearly 13.7 million dwellings. Almost 70.5 thous. strata were established, while the sample size in particular strata ranged from nearly 6% to over 49%.

#### **1.4. Combining data obtained from different sources**

The diversity of survey methods and techniques of data acquisition for the purpose of the National Population and Housing Census 2011 triggered the necessity to combine data coming from various registers and information systems with sample survey results and census results obtained through the Internet. A wide range of data was collected in the sample survey from over 8 million persons, either residing or staying in approx. 2 744 thous. dwellings sampled. A vast majority of data in this survey was gathered through direct interviews conducted by enumerators with respondents. The percentage of persons who completed the long form in the Internet self-census amounted to less than 2% whereas the share of persons participating in the self-census and completing the short form exceeded 10%. Telephone interviews were mainly conducted in the case of incomplete responses provided through questionnaires, in order to obtain responses to the questions omitted, or clarifications of any inaccuracies and inconsistencies in the data collected.

Considering the solutions applied in the sample and full survey, it was possible for one person to be enumerated more than once. Therefore, at the processing stage of census results, it was necessary to select the source of data and “the best” individual questionnaire in which the widest range of data was acquired. At the same time, it was assumed that the questionnaires considered were complete and closed, which means that for each individual questionnaire all information was obtained, resulting from following all skip patterns in the census form. The procedures of questionnaire selection were conducted upon verifying and determining the

completeness of respondent coverage, based on the information included in the address and housing register, updated after the census and dwelling deduplication. In the case of the sample survey, this process entailed selecting a set of “the best” individual questionnaires for each dwelling, referred to as “the individual package” (understood as a set of individual questionnaires related to a given dwelling and enumerated in this dwelling). In turn, short self-census forms were deduplicated individually, based on a doubled or multiplied identification number (PESEL), both within the same dwelling and between various dwellings. The deduplication procedure of these questionnaires was more complicated in the absence of the PESEL number which constitutes a unique personal identifier.

Information obtained from registers and information systems was subject to special validation which covered normalisation, conversion, as well as subjective, objective and spatial synchronisation. Data coming from registers and information systems might be the source of non-sampling errors, given both the lack of integrity between such databases and unidentified differences in definitions. A special approach had to be taken with respect to those situations where errors occurred in the PESEL number, and when there was inconsistency between the PESEL number and the person’s date of birth. It should also be borne in mind that the registers and information systems which were used in the 2011 Census were established on the basis of legal provisions adopted for very different purposes than statistics. Each entity included in any such register or information system may not be removed from the database unless the circumstances defined in the provisions of law occur, and unless an appropriate document confirming the change of the legal status of any such entity is delivered.

Data to be used in the 2011 Census was collected from administrative sources upon verification based on the information collected from respondents as part of the Internet self-census (through the short form) and the sample survey (through the long form). Information obtained from registers and information systems was altogether updated for the purpose of the full census for approx. 32% of persons.

**The census database established required the application of two approaches while processing the 2011 Census results.**

**The first approach** was applied with a view to determining the population status and characteristics according to demographic characteristics (sex and age) and place of residence. The completeness of this data for the entire population in the country was safeguarded by means of updated data contained in registers and information systems. This implies that the population statuses and demographic structures, such as sex and age, could be determined for any small administrative division unit of the country, and for the territorial (localities), geographic (coordinates x, y) and statistical (census areas, statistical regions and statistical localities) profiles.

**The second approach** was applied with respect to most topics surveyed in the census for which the data included in registers and information systems was insufficient. For those topics the census results were derived by the generalisation of the data collected in the 20% sample (in national terms) of persons residing or staying in the dwellings sampled to the total population. An unprecedented size of the random dwelling sample used in the 2011 Census, which was selected from over 70 thous. strata comprising all gminas in the country, ensured

representative results from the powiat level upwards. The 2011 Census results derived from the sample survey should always be presented by indicating the precision error. The information gathered for 20% of the population comprises a range of data which is broad in subjective terms, and which, after being generalised to the entire population, provides detailed demographic, social and economic characteristics of the population, households and families.

### 1.5. Generalising the sample survey results

While processing the sample survey results, two generalisation (weighing) stages of the results were distinguished:

- Stage I – weighing through adjusted “fraction” weights, constituting the opposite of the sampling fraction, for housing units of the survey,

Stage II – weighing through calibrated weights for survey units (persons) The primary weights were established as the opposite of the sampling fraction for nearly 70.5 thous. strata. It should be borne in mind that the aim of stratification was to identify as homogenous groups of sampling units as possible. Weights applied within each stratum were identical. The initial fraction weights had to be adjusted, given that 13.7% of housing questionnaires were missing from the sample survey. **The adjusted weights established during Stage I are used to generalise the census results concerning dwellings.** However, separate weights were derived for single-family buildings which comprise one or two dwellings, as well as for households and families. Such weights were calculated on a secondary basis through calibrated weights for persons forming households.

Given the need to integrate the representative survey results with the full census (in terms of underlying variables such as sex, age and place of residence at the powiat level, including the distinction into its urban and rural parts), it was indispensable to derive calibrated weights for particular persons.

**Calibration is a method entailing an appropriate weights adjustment** so as to compensate for information losses resulting from non-response. The purpose of calibration method applied for the 2011 census was to adjust the structures of sex and age acquired in the sampling survey to the structures of population according to sex and age, as established in the full survey. Its results constituted a reference population. Weights calibration was performed using the CALMAR (*Calibration on Margins*) programme by representatives of the academics from the Poznań University of Economics.<sup>4</sup>

**The calibrated weights are used to generalise the sample survey results concerning the population statuses and structures.** They allow recreating the structure by sex and age for the largest five cities, i.e. Warsaw, Kraków, Łódź, Poznań and Wrocław. For Warsaw this can

---

<sup>4</sup> A working paper entitled: “Raport z opisem wyników z zakresu możliwości wykorzystania kalibracji na potrzeby korygowania wag w złotym rekordzie” (*A report with the description of results regarding the possibilities of applying calibration for the purpose of adjusting weights in the golden record*), prepared under the supervision of Prof. Jan Paradysz, PhD hab.

be done for sex and the age of 0,1,...,99, 100 years or more, whereas for the remaining four cities – only until the age of 85 years or more.

It is possible to recreate the population structures by age for voivodships and for the whole country, obviously taking into consideration such variables as sex and locality class (including the distinction into urban and rural parts). However, the last age group should be established as 85 years or more, given the lack of representatives in the older age groups for some voivodships in the sample survey. Due to weights calibration, the census results generated from the full survey and generalised from the representative one are identical as regards the population structure by sex, age (with a certain limitation concerning age groups) and locality class.

### **The algorithm of establishing weights for dwellings in the sample survey**

With the aim of obtaining accurate generalisations, it was necessary to adjust the initial weights resulting from the sampling scheme. The adjustment took into consideration both non-response and information concerning the reasons for failures to obtain data from some of the surveyed units. As dwellings served as the sampling units, the adjustment procedure made use of the information assigned to the variable “reasons for not completing the questionnaire”, as part of which different causes were coded for different dwellings sampled.

Within each stratum defined at the dwelling sampling stage, the weight adjustment was determined on the basis of the sampling fraction defined for a given stratum. The adjustment entailed using the adjustor calculated as a quotient of the number of dwellings which should have been surveyed to the number of dwellings which were actually surveyed (the opposite of this quotient can be interpreted as the indicator of completeness). Within the selected sample of 2 684.2 thous. dwellings, completed forms were collected from 2 317.7 thous. dwellings. The instances treated as falling beyond the scope of the population surveyed, i.e. where there was no dwelling, or where the dwelling was unoccupied, subject to inheritance proceedings, intended for repair, or not inhabited yet, constituted 140.4 thous., i.e. approx. 5.2% of the sample. Such cases occurring in the sample were considered sampling frame errors.

The adjusted weight assigned to the dwellings surveyed in a given stratum was a quotient of the weight adjusted to the sampling (fraction) weight. In the case of strata for which the adjustment calculated was too high, exceeding the agreed threshold (adopted at level 3), the procedure of “combining” adjacent strata was applied and the adjustment calculations were repeated until the limitation assumed was complied with. Such a procedure was used to prevent an excessive diversification of weights, which increases estimators variance and has a negative impact on the precision of estimation.

Those dwellings which were included in the sample after the sampling stage had been completed were approached separately. Within the group of 105.4 thous. dwellings additionally surveyed, the survey was conducted in 45.3 thous. dwellings, 43.1 thous. of

which turned out to be outside the population surveyed. For this part of the sample it was assumed that the sampling weights equalled 1 and the described adjustment algorithm was applied at the level of random strata defined as powiats.

The weights described above can be used to derive global values (e.g. the total number of dwellings, occupied dwellings or substandard dwellings) and to estimate dwelling characteristics (e.g. dwellings by size).

### **The imputation of the missing data**

The imputation method was used in the 2011 Census with a view to determining the data missing from the responses provided. This was the so-called deduction-based imputation, using the correlations between variables. In most cases it involved item imputation, conducted by taking the deterministic approach. It was also applied in the case of missing responses in the sample survey forms, based on the variable values available in registers or information systems. Such procedures could only be performed on the condition of full compliance in terms of definitions of the variables occurring in both data sources. In infrequent cases, imputation of the entire units surveyed was made, based on the stochastic approach.

## **1.6. The global values estimation precision**

Global values for individual variables constitute the principal parameters calculated on the basis of data gathered in the sample survey as part of NSP 2011. Most of the variables surveyed concern the population of persons, and a smaller number of variables describe dwellings and buildings. The estimation method and the estimation quality assessment regarding the population characteristics are described below.<sup>5</sup>

The global value estimator for particular individual characteristics at the country, voivodship or powiat level corresponds to a weighed sum which is calculated as follows:

$$\hat{Y} = \sum_{h,i,j} w_{hij} y_{hij}$$

where  $h$  indicates the strata resulting from the dwelling sampling scheme, defined through the combination of variables, such as the gmina code before the census interview and the strata number, whereas

$i$  – means the dwelling code, i.e. the sampling unit,

$j$  – means the person included as a dwelling occupant,

$w_{hij}$  – means the weight assigned to a given person from a give dwelling; this is the value recorded in the variable *weight\_per\_population\_first\_person*, i.e. after calibration, and

$y_{hij}$  – means the value of the individual characteristics analysed, e.g., 1 – if a person has attained the level of higher education, 0 – otherwise. For specific result tables it may prove necessary to design a special set of new characteristics, based on the ones existing within the output data, e.g. to analyse the populations by age group, dummy variables should be defined for specific age intervals. If deeper grouping levels are necessary, additional variables should be defined as appropriate (e.g. taking into consideration the distinction into urban and rural areas, or sex).

---

<sup>5</sup> A working paper prepared by Robert Wieczorkowski, PhD.

**To assess the quality of the global value estimator, the so-called estimation precision indicator is established.**

It enables the output table users to properly assess the data included. It was assumed that the precision indicators presented will be defined as relative standard errors (in other words, coefficients of variations, referred to as CV, and expressed in %), based on the following formula:

$$CV = 100 * \frac{\sqrt{Var(\hat{Y})}}{\hat{Y}}$$

where  $Var$  reflects the assessment of the estimator variance, calculated in the following way:

$$Var(\hat{Y}) = \sum_h \frac{n_h(1 - n_h/N_h)}{n_h - 1} \sum_i \left( \sum_j w_{hij} y_{hij} - \frac{\sum_i \sum_j w_{hij} y_{hij}}{n_h} \right)^2$$

Value  $n_h$  indicates the number of dwellings in a given stratum of the sample drawn, whereas  $N_h$  stands for the number of dwellings from the upper population stratum; value  $N_h$  may be retrieved by adding up the adjusted weights which result from sample selection for the purpose of the sample survey.

**The precision estimation indicators for the global values of the underlying variables are presented in the precision tables, to be found at the end of the tabular annex in each publication with the results of the 2011 Census..**

An exemplary interpretation of the precision indicator for the global value, estimated on the basis of the sample survey, reflecting the number of persons in Poland holding a scientific degree of at least a PhD is as follows: CV=0.85% means that the actual value of the parameter estimated equals  $153.8 \pm 1.3$  thous. (this is referred to as point estimation).

CV may also be used to determine the confidence interval at which, with certain probability (i.e. the confidence level of, e.g., 95%), the actual value of the parameter estimated is covered. For instance, for the confidence level of 95% the borders of such an interval are as follows:  $\hat{Y} \pm 1.96 * \hat{Y} * CV / 100$ . If the estimated number of persons holding a scientific degree of at least a PhD amounted to 153.8 thous., and the precision indicator for this characteristic equalled CV=0.85%, then it could be determined, by specifying the borders of the confidence interval, that with the probability of 0.95 the range of values from 151.2 thous. to 156.4 thous. covered the actual number of such persons.

Here is another example. The precision indicator for the same variable, i.e. the number of persons holding a scientific degree of at least a PhD, in lubuskie voivodship amounted to 5.3%. If the estimated number of persons holding a scientific degree of at least a PhD amounted to 1770, and the precision indicator for this characteristic was CV=5.3%, then it could be determined, by specifying the borders of the confidence interval, that with the probability of 0.95 the range of values from 1586 thous. to 1954 thous. covered the actual number of such persons.

Such calculations indicate that in rare cases (involving smaller numbers), which result from a considerable scattering of units, or from an uneven distribution of variables, the global values of such variables, estimated on the basis of the sample survey results, are burdened with

higher sampling error. In the first example, the value obtained in the sample census differed from the actual value by  $\pm 1.7\%$ , whereas in the case of the same variable for lubuskie voivodship this difference grew to  $\pm 10.4\%$ . With the precision indicator of  $CV=10\%$ , the generalisation error of the variable coming from the sample survey, with the probability of 0.95, is actually doubled and amounts to exactly  $\pm 19.6\%$ . Therefore, the sample survey results disseminated at the confidence interval of 0.95, where the precision indicator exceeds 10% (CV), should be interpreted with caution, whereas the results for which the precision indicators exceed 20% of the CV value should be disseminated in an aggregated form.