

Integration of Agricultural Census and Population Census Data

Eva Laczka*

Hungarian Central Statistical Office, Budapest, Hungary

e-mail: eva.laczka@ksh.hu

In Hungary a general agricultural census (AC) was carried out in 2000, followed in 2001 by the population and housing census. The two censuses had been designed separately. Originally the Hungarian Central Statistical Office (HCSO) did not plan the joint analysis of the data of the two censuses. Following the censuses users and researchers expressed the view that linking the data of the two databases would represent a value-added in the use of the data and the joint utilization of the databases of the two censuses was examined. The databases were matched and the aggregated handling of the information increased the potential for analysing both censuses and allowed further, more sophisticated investigations. By means of the databases of the two censuses, the first opportunity arose for matching the discrete data of the surveys. The precondition of the matching of the data was the conformity of the respective metadata of the two operations. „Private holding” and „dwelling-household” were the categories applicable as the smallest unit for the matching. The links between the private holdings and the households could be based on the identity of the persons living in the dwelling. The matching of the data required the use of individual identity codes. With the matching process used a joint database of the agricultural and population censuses was set up providing new approaches for gender disaggregated analysis. By using the linked database, the HCSO issued a series of publications on the households living in agricultural private holdings in the countryside. This presentation describes the method of matching the databases of the two censuses.

Keywords: censuses, integration of censuses, matching method

1. Introduction

The Hungarian Central Statistical Office (HCSO) carried out in April 2000 a general agricultural census (AC 2000) and in February 2001 a population and housing census. Although at the time of the planning and conduct of the two censuses there were no plans at all to link the execution of the two censuses and the two datasets, due to the proximity in time of the two operations, statisticians decided in 2003-2004 to apply a new type of solution in order to explore the maximum potential offered by the two censuses.

The results of the work were published by the HCSO in a series of publications entitled “Private holdings and their population”. The publications analyze private holdings and the situation of agrarian population in 7 planning-statistical regions. Their content will be outlined at the end of the paper.

2 What gave the idea?

Agriculture has always played an important role in the Hungarian economy which is also due to the excellent natural conditions of the country. The experience of several decades confirms that – traditionally – nearly half of the population of the country deals to a smaller or greater extent with agricultural activity.

In 2000 during the conduct of the AC 2000 enumerators visited 2.1 million households (which represented at the turn of the century nearly half of Hungarian households). According to the results of the census nearly half of the households visited carried out agricultural activity, among them 960 thousand households were considered

agricultural holdings from a statistical point of view. Hungarian traditions are well characterized by the fact that besides the 960 thousand households mentioned, further 835 thousand households carried out agricultural activity without being farms (under the threshold), and only 300 thousand of the households visited did not perform agricultural activity.

3. Main characteristics of households

These data already reflect the specific structure of Hungarian agriculture: it is characterized by a great number of small “holdings”. Nearly half of the households (the so-called individual farms) produces exclusively for own consumption. In the economic sense these cannot even be considered as “farms”, their production is more based on traditions and social reasons. This does not mean however that we can neglect the output of the production units not considered as farms in the economic sense, as they represent 10-15 per cent of the output of Hungarian agriculture. At the turn of the century the proportion of market oriented private farms (households) was only about 10 per cent. Besides the two types of farms mentioned, there is a significant number of private farms producing both for own consumption and for the market.

The 2001 population census was the 14th census in the history of Hungarian statistics. The enumerators collected data about 3.9 million households, dwellings. The questionnaires of the population census contained questions about the gender and age composition of households, they gave detailed information about the qualifications, education, employment of the members of the households and provided a picture about the demographic and social conditions of the Hungarian population, coupled with questions about dwelling conditions.

4. The idea

On the basis of the above mentioned results, statisticians decided to analyze more in detail the social conditions of the agrarian population, for which they used as a tool allowing a complex analysis of the two datasets the – posterior – linking of the databases of the two censuses. We statisticians thought that an integrated database would make possible a detailed examination of the social stratification of the people living in individual farms. We analyzed the social situation of the Hungarian agrarian population using important indicators, like the employment situation of the farmers and the persons belonging to the farm-household, the type of activity, the professional and other educational qualifications of farmers, their place and role in the social hierarchy. Besides the type of employment, we also investigated whether farmers performed their activity as self-employed persons or as employees, if they carried out intellectual or physical activity, and what was their income situation. The types of settlements, dwellings where individual farmers lived, their living conditions were also examined. The investigation and analysis of these issues required the creation of an integrated database.

As during the design phase of the two censuses it was not planned to link the databases, the task to be solved was not easy. The questionnaires of the AC 2000 contained the names and addresses of the farmers (according to the Hungarian data protection rules these could be used exclusively for statistical purposes), while in the case of the persons living in the households these information were missing from the population census form. Thus a totally new solution had to be elaborated for linking the databases.

The Hungarian statistical service has never had before such wealthy databases on the population collected within such a short lapse of time. The joint utilisation of the

databases, the aggregated handling of the information expanded the potential for analysing both censuses and by merging and matching the databases opened the door for further, more sophisticated inquiries. Though the population census provides basic data for social statistics, a part of the information collected is less needed by the statisticians dealing with industrial, economic statistics. On the other hand the agricultural census, one of the most important data collections in economic statistics, collects information which is significant for the better understanding of certain social processes, e.g. for more in-depth analysis of the motives of demographic developments.

Besides the practical aspects, the joint utilisation of the two datasets was supported by the fact that none of the independent programmes covered the areas and the information which had been dealt in detail by the other. It is important to mention that both censuses were full-scale operations, as a consequence discrete data are available. The latter facilitates the aggregation of the data according to any optional territorial unit, criteria. The former grouping, analysis requires the highly developed IT infrastructure of our days, permitting the aggregation of the data of the two data collections.¹

By means of the databases of the two censuses, the first opportunity arose for matching the discrete data of the surveys. This is a new development not only in the history of the Hungarian censuses but it has broken a new ground in international practices as well. Besides the impetus that was given to commence the job by the accession of the country to the European Union, an opportunity was given for the description of the demographic and housing characteristics of the population working in agricultural private holdings. The results of this exercise, with special regard to the territorial presentation, serve as a proper platform for future surveys.

It was clarified already at the beginning of the works that the precondition for matching the data of the two censuses was the conformity of the respective (common) metadata of the two operations. The aim was to approximate the different concepts referring to similar groupings that had been applied in the two censuses respectively. By investigating the issue it turned out that “private holding” and the “household” are the categories practicable as the smallest unit for the matching. Consequently, the experts decided to utilise the following categories: the private holdings² from the agricultural census and the households, more exactly the dwelling-households.³ There was a general agreement that, despite the fact that the two censuses cover different aggregations, the categories used more or less refer to the same population groups. Moreover, thanks to the high frequency of the categories in the two censuses, there is a rather large common mass and only a limited circle is represented only in one of them.

The precondition of creating a common database was the univocal matching of the elements of the two surveys following the separation of the data with the identifiers defined above. Further analysis of the common database was based on the elements (private holdings and households respectively) marked with the same

¹ Besides meeting the technological requirements, the protection of personal data should be secure. According to the respective law, the matching of the data is allowed only for statistical purposes.

² The category „private holding (farm)” covers the families, communities of persons living in kinship or not, using a common dwelling, performing agricultural activities as private entrepreneurs, self-employed, family helpers.

³ The category „dwelling-household” means the total number of people living in a dwelling. It is to be stressed that in this paper „household” – unlike the category used in Hungarian population censuses – refers to the “dwelling-household”.

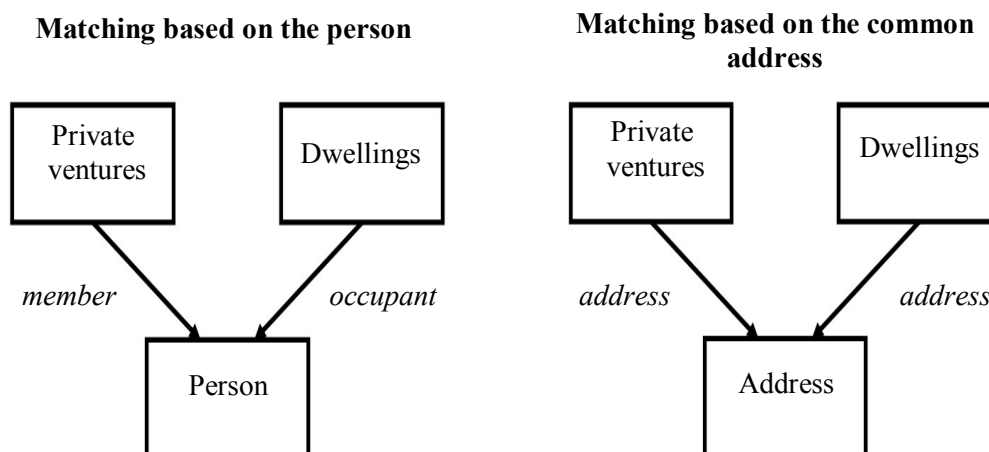
identifier. The most suitable tool for the matching of the former elements was the address of the enumeration.

The Law XLVI of 1993 on statistics serves as legal background for matching the two databases.

The aim of the project was to build up with the help of the appropriate computer programs a common database suitable for transferring/accepting elementary information from both censuses. The content of the database had been fixed upon the requirements of the planned investigation. By using the linked database the Hungarian Central Statistical Office issued a series of publications on the population living in agricultural private holdings in the countryside (i.e. outside of Budapest).

5. Matching method

The aim of matching the two databases was to study the relationship between the categories „private holding” and „household” used by the agricultural census and the population census, respectively. The correspondences are based partly on overlapping persons, partly on the identification of the real estates (addresses) of the given categories.



The units of observation were private holdings in the AC, while dwellings and persons in the population and housing census. The agricultural census recorded some information (e.g. gender, age) also on persons other than employees working in the private holding. Based on these records it was possible to estimate a link between members of the private holding and occupants of the dwelling. The only common base of connecting the given real estates was the address. The links between the private holding and the household can be based on the identity of the persons living in the dwelling. According to the theory of the databases the matching of the data requires the use of the individual identity codes. The individual identity code is a (contingently complex) parameter which defines the elements (characteristics) of a mass univocally. The links are represented by the couples formed by the identical codes of the occurrences.

The steps of the matching process were the following:

- nearly 90 percent of the 1 million agricultural private holdings have been identified by using a special programme for matching the elements of the two databases of the addresses;

- the remaining, “unmatched” AC addresses were recognised with the help of an interactive programme which resulted in a further 7 percent of matched addresses; the process lasted for around 10 months;
- the identification of the persons was performed later on, by another programme: 78 percent of the members of agricultural private holdings could be “paired” with persons in the population census.

Based on the above described procedure a joint database of the agricultural and population census was set up providing new approaches for the gender disaggregated analysis.

6. The result

The establishment of the joint database made possible the publication of seven volumes entitled “Private holdings and their population” covering the seven planning-statistical regions of Hungary. Each volume follows the same content pattern:

Introduction

Country characteristics

- Private farming
- Main characteristics of county population and dwellings
- Main characteristics of the population and dwelling conditions in private farms with different production aim

Characteristics of the region

- Farm structure of private holdings
- Population, demographic characteristics of private holdings
- Employment characteristics of the persons living in private holdings
- Household composition of private holdings
- Dwelling conditions of persons living in private holdings

Small area characteristics in the region

- Small area farm structure
- Main demographic characteristics of the population of private holdings in small areas
- Small area characteristics of economic activity
- Household composition in small areas
- Dwelling conditions in small areas

The purpose of the publications was not only to serve as a basis for analysis, but statisticians wanted to raise awareness about and call the attention to the fact that the linking of different databases could result in obtaining more complex information allowing a better comprehension of economic and social processes, as well as taking advantage of the synergies between different statistical domains.

7. Conclusion

The initiative to link the databases of the two censuses proved to be very successful. As a result of matching the two datasets researchers, users had at their disposal a unique tool to make disaggregated, complex and multidimensional analysis of the characteristics of private holdings and their population. They claimed the usefulness of such initiative in case of future agricultural and population censuses as well.

Statisticians were also in favour of linking the two different statistical databases, preferably already in the design phase of the two big full scale operations. This was our intention in case of the latest round of censuses. Unfortunately due to some data protection concerns the decision was finally not taken prior to the censuses. Thus the solution used 10 years ago could be followed again.

References

“Magyarország mezőgazdasága a 2000. évben- Összefoglaló adatok”, *Központi Statisztikai Hivatal, 2002* (“The agriculture of Hungary in 2000, Summary data”, Hungarian Central Statistical Office)

“Az egyéni agrárgazdaságok és népességük – Közép-Magyarország” *Központi Statisztikai Hivatal, 2004* (“Private holdings and their population – Central Hungary”, Hungarian Central Statistical Office, 2001)

“Az egyéni agrárgazdaságok és népességük – Közép-Dunántúl”, *Központi Statisztikai Hivatal, 2004* (“Private holdings and their population – Central Transdanubia”, Hungarian Central Statistical Office, 2004)

“Az egyéni gazdaságok és népességük – Nyugat-Dunántúl”, *Központi Statisztikai Hivatal, 2004* (“Private holdings and their population – Western Transdanubia”, Hungarian Central Statistical Office, 2004)

“Az egyéni gazdaságok és népességük – Dél-Dunántúl”, *Központi Statisztikai Hivatal, 2004* (“Private holdings and their population – Southern Transdanubia”, Hungarian Central Statistical Office, 2004)

“Az egyéni agrárgazdaságok és népességük – Észak-Magyarország”, *Központi Statisztikai Hivatal, 2004* (“Private holdings and their population – Northern Hungary”, Hungarian Central Statistical Office, 2004)

“Az egyéni gazdaságok és népességük – Észak-Alföld”, *Központi Statisztikai Hivatal, 2004* (“Private holdings and their population – Northern Great Plain”, Hungarian Central Statistical Office, 2004)

“Az egyéni agrárgazdaságok és népességük – Dél-Alföld”, *Központi Statisztikai Hivatal, 2004*

“Population Census 2001”, *Hungarian Central Statistical Office* (“Private holdings and their population – Southern Great Plain”, Hungarian Central Statistical Office, 2004)

“Az egyéni gazdaságokban élők társadalmi rétegződése”, *Központi Statisztikai Hivatal, 2006* (Social stratification of the population of private holdings”, Hungarian Central Statistical Office, 2006)