SUMMARY

CENSUSES AND PUS

The population census is one of the most important stores of information in Swiss official statistics. A great set of data in all senses of the word, it records information on the demographic, professional and social situations of the population across the country, and can also be used to analyse information on households and housing conditions. Furthermore, it is repeated every 10 years, thereby providing a clear picture of changes taking place in Switzerland. This is why this information should be easily accessible to researchers and anyone else with an interest.

The same qualities for which a census is prized can make it difficult to analyse – take, for example, the size of its dataset and the vast scope it covers. Even with modern computer technology, a dataset of several million individuals still makes for complex processing. Then there is the even greater problem of how to keep the information confidential. Given the comprehensive nature of censuses, it would be feasible for their data to be combined until someone is identified – something that is unacceptable both ethically and legally. This explains the decision of the Swiss Federal Office of Statistics (OFS) to draw a random representative Public Use Sample (PUS) of 5% of the population that ensures the preservation of anonymity[1] while allowing data to be available freely without prior restrictions.

The information described here goes beyond the 2000 census and dates back to that of 1970. This means that professional, regional, and other changes in Switzerland during this period can be analysed. Comparing several censuses requires a specific approach to harmonise the data; after all a census, just like any survey, is a reflection of the age in which it takes place. In 1970 computer-related jobs, for example, were not the same as they are today, hence the need to adapt their names. Quite apart from any technical issues, then, the question of how to compare data collected at different times deserves careful consideration.

This last remark also clearly highlights the importance of the documentation: it must reflect all the aspects regulating the compilation of the data, and also take into account the reasons for preparing the data and the working methods of potential users. This data description is based mainly on a tremendous effort by the OFS of compiling technical documentation as well as on the original census information found in the numerous OFS publications. We have nonetheless tried to present a description that is simple and accessible, while including information from several publications.

WHO WOULD USE PUS AND WHY

In sum, the purpose of the work presented here is to allow any interested party to analyse the census data in the context of a "scale model", especially with a view to:

- teaching, for example, statistical analysis courses or other research courses or lab work, in order to introduce students to census analysis and to increase their awareness of the subject;

[1] Apart from making the data anonymous in accordance with the law on the protection of information – that is, by removing all indications relating to names and addresses – a second process was used to eliminate possible identification by cross-checking. Of course, this does not preclude the use of the original data, as long as the rules of deontology and confidentiality are upheld and guaranteed by the signing of a specific contract.

- research or preparatory investigation for more in-depth research, such as exploratory or preparatory laboratory analysis to encourage all research to be initiated with a preliminary study based on public statistics;
- research preparation requiring knowledge of certain population traits, or of quotas, estimates of the size of specific groups, classification preparation, aggregation assays, etc.;
- accessing contextual information in order to supplement data collected during other research activities, which implies creating a link on "geographical levels" and the aggregation options they represent in the documentation.

The target audience includes any interested party irrespective of activity, but in particular:

- students within the context of courses and seminars;
- doctorate candidates and scientific researchers at universities, who would be able to familiarise themselves with official statistics while learning to formulate the questions to be asked more specifically and to use the data necessary for their answers;
- those employed in private institutions and state administrations, who will better realise the potential of census-taking and benefit from the easy access to global data;
- foreign researchers, especially if they already work on PUS in their own countries.

One of the major benefits of PUS is paradoxically not in the data we are presenting here. Since the information was restricted to ensure anonymity, certain analyses are prohibited or restricted. However, by showing the wealth of information that exists and revealing the options afforded by censuses, PUS facilitate access to census data: it is easier to make a request to the OFS when one knows which variables are available, and then to pursue the analysis on the entire database.

WHICH INFORMATION SHOULD BE MADE AVAILABLE

A census includes several sets of data – on people, households, housing and buildings. To simplify its use, we have decided to consider here only the file on people, to which we have added information on the households, buildings and housing. In technical jargon, this would be referred to as a Public Use Sample of people and not as a Public Use Sample of households.

For each census, the sample is obtained by a simple random drawing of people from the complete dataset to obtain a 5% sample of the population. As for the content, we did our best to use all the available information and only eliminate where necessary to prevent a respondent from being identified. In this respect, we systematically chose to limit the detail of the answers rather than select a subset of variables.

The philosophy behind this project is therefore that of a "loss leader": by revealing all the data that is available, we hope that users will be tempted to make use of the entire census, even if this means committing themselves contractually to preserving anonymity.

In concrete terms, this product comprises three kinds of content: information on the censuses in Switzerland and especially the data collected in this context, some indications on the requirements of data protection and the means to meet them, and finally comprehensive documentation on the available variables.