

Data Processing of the 1999 Vietnam Population and Housing Census

*Prepared for UNSD-UNESCAP Regional Workshop on Census Data Processing: Contemporary technologies for data capture, methodology and practice of data editing, documentation and archiving
Bangkok, Thailand, 15-19 September, 2008*

Mai Van Cam
General Statistics Office
Vietnam

This paper divided by three parts as follows, Part I summarizes recent population censuses in Vietnam; Part II presents Data Processing of the 1999 Vietnam Population and Housing Census; and Part III describe some strategies would be applied in the 2009 census.

I. Population Censuses in Vietnam

Viet Nam has a long record of census taking dating back many centuries. However, most of the early censuses were really no more than population counts designed to keep track of persons who were required to pay taxes or who might be needed to fight in local wars. Thus they were sporadic and sought few details. The first real census of an independent Vietnam was conducted late in 1979. Given the resources and technical skills available at that time, the census provided surprisingly good benchmark data as a springboard for national development.

The first census which could really be thought of as a modern census, introducing internationally recognized census concepts, design features and

processing, was conducted in April 1989¹. It was felt by the many people and organizations who participated in this census that the coverage of persons resident in Vietnam was near complete and the results were of very high quality. For the past decade this census has provided a rich source of demographic, social and economic data for a wide range of users.

The third census was conducted in April 1999. As will be seen, many of the features of the 1989 census were incorporated into the design and conduct of the 1999 census. In addition, however, the 1999 census added new questions and extended its scope in some areas to provide even more comprehensive data. The two censuses together will provide a rich source of data to analyze the current situation and key trends over the past ten years.

II. Data Processing of the 1999 census

2.1 Selection of software

Questionnaires used in the main 1999 census² were designed in booklet format, where separate columns were provided across each page to record information for each person. While these booklets proved very useful in recording and checking information, they presented some problems in data processing and in selecting and designing software for data entry. In late 1998, the Central Data Processing Centre (CDPC) conducted test runs to develop its data processing system, using records from the pilot census. Three different applications were tested for data entry (adapting standard packages known as IMPS, ISSA and FOXPRO). From these tests it was decided that the version based on ISSA was the best-suited to the Viet Nam census.

¹ Census date was 1st April 1989.

² Census date was 1st April, 1999.

2.2 Distribution of data processing facilities

Data entry and on-line editing facilities were provided at 9 centers (CDPC, Hanoi, Nam Dinh, Da Nang, Khanh Hoa, Binh Thuan, Ho Chi Minh City, An Giang and Can Tho). Each centre established a computer network for census data processing with a server HP LH3 and 12 to 53 PCs. For the entire country, 240 PCs were used for processing and 10 servers (two in CDPC). A total of 450 data entry operators were engaged to work in two shifts. The network in each regional computer centres is connected to the CDPC network through a telephone dial-up system. Thus, as data entry was completed in each province, data files were transferred to the CDPC server.

2.3 The data entry, editing and tabulations operations

The main strategy for data entry is to "photocopy" all information on the questionnaires to the computer, that is to change recorded information as little as possible. To do this keyers were not permitted to modify or correct recorded information. However, there were a number of checks provided to ensure the keyed data were valid, but most of these checks were in the form of warning messages to control for key stroke errors and column shifting. Data entry was completed for the 3 percent sample within about two months (working July and August 1999). All editing operations were undertaken at the CDPC. Tables were produced progressively from September 1999, using the CENTS module.

A number of consistency checks were carried out (the edits were specified in CONCOR) and records in error were edited on-line by special editors. As soon as data entry for a district was completed, a listing of inconsistencies was printed out for checking and, correction and updating of files.

2.4 System management and control

A system of management and control was developed by Central Data Processing Center (using a VISUAL FOXPRO application) to help

managers to monitor the processing. All stages in the cleaning of data are monitored under this system, from the receipt of the questionnaires, to data entry, verification, checking, listing of inconsistencies, data correction, combining EA data files into higher geographic levels, production of frequency tables, and data backup. The system also provides the mechanism to validate the geographic identification of keyed data, to avoid duplication or omission of EAs. For managers, the system generates different kinds of reports, for example, to keep track of the status of each EA, to calculate the quantity and quality of work of the data entry operators and print salaries due, or to provide frequencies of imputed values to subject matter specialists to ensure the rules were properly applied.

2.5 Data Dissemination

The requirement of statistics is increasing rapidly from the previous census. There are urgent needs of providing socio economic statistics for the whole country, and in development of Information Technology (IT) i.e. for storage, communication and exploitation of data. Policy makers, planners, researchers, international organizations and users in different economic sectors at every level wish to use data from the population and housing census. The data source is the base for estimates on socio economic achievements over the past ten years. It helps us to understand the impacts of migration to urban areas, and changes for people in the economic, socio and demography aspects.

The success of the 1989 population census is a good condition for the users of information. This is the first time that results from both 1989 and 1999 censuses can be analyzed together. The 1999 Census has mainly applied the same strategy as the 1989 census design including the questions on the long form (for the complete census) and short form (sample census) as well as main information on housing. However, contents of the 1999 Census have been remarkable expanded when compared to 1989 Census: many new questions have been added into the long questionnaire form. The scope of

the housing census has been expanded and this is the first time a complete housing census has been conducted in Vietnam. Many questions have been added were some questions are for every household in the whole country while other questions were intended for households in rural areas only.

Contents of the census are rich and complex but the budget for it is rather limited when comparing with some international standards. The Vietnam Government has taking care and supporting the most necessary finance requirements for the Census. The Census has also received positive support of international organizations. UNFPA was in the beginning concerned about their ability to directly give financial support to the Census, but in fact they have given both financial and technical support including new equipment. They also took an important role in the cooperation with General Statistics Office to find out some resources for the Census such as from UNDP, Governments of Holland, Denmark and Australia. Most of the resources have been used for training, equipment for data processing as well as analysis and dissemination of results from the Census.

The ability for providing of the first results only one year after completing of the Census is a very nice effort to be able to have the data dissemination to users timely. Although the sample results are reliable in general, it is careful when explaining and using the census results because of some sampling errors. It is the same in the case of some issued figures in analysis have been rounded until several thousands units or only are represented on graphic/map while sampling errors is included in small value data. Of course the results from the complete census will have no sampling errors. The indicators on fertility and mortality, because of being collected from sample census, are allowed to tabulate and expanded to the provincial level.

Detailed reports on the complete census will be completed at the end of year 2000. Apart from some reports at national level, the reports raised from research requirement at province/district level will also be compiled. The subjects on housing, migration, labour and employments as well as

demography have been taken into some plans. How to provide good information services to users is also an important thing, which is, discussed in detail between producers and users.

In the 1999, besides the traditional paper books of tables, analysis reports of the census results, the following electronic media were made:

i. The CD ROM for data and result of the 3% sample is produced and can be seen as one of the electronic product to meet the information needs of users.

There are two main modules in the CD ROM:

First module: Keep micro data of 3% sample (from household and individual questionnaires) in IMPS format and layout about each item can be found in the data dictionary.

In addition to the data file, the CD ROM also provides some modules of IMPS (Integrated Microcomputer Processing System) and its applications as tool for producing tables and thematic map.

In IMPS, Crosstab is a module that allows users to make their own statistics table. It is located on the CD ROM included micro data and dictionary. MapView is a module that helps users to create electronic maps on population information. It can be seen a simple electronic population ATLAS.

Second module: Contains tools for converting the original raw data files from the IMPS environment into more common file formats such as delimited text for imports to Excel, SPSS, and Microsoft Access.

ii. CD ROM as the same as above mentioned for each provinces. There are 64 CD ROM for 64 provinces.

iii. CR ROM “Census PopMap application”: This provides a database with 232 indicators of the 1999 census at level of whole country, province, districts and some basic mapping classes such as administrative units, transport lines etc.

iv. Socioeconomic atlas of Vietnam based on the 1999 Population and Housing Census titled “A depiction of the 1999 Population and Housing Census was created. This publication then also have produced in CD ROM for easy use.

III. Preparation of General Statistical Office for the 2009 census

It is planned that the next population and housing census of Vietnam will be conducted at the time point of 0 hours of 1 April 2009. Some strategies would be applied in the 2009 census as follows:

i). To use sample survey technology to expand the census content and to economize the census expenditure.

In order to improve the effectiveness of the census design, it is intended to apply the strategy of two-phase approach for questionnaire design. A comprehensive questionnaire (short form) contains only some core questions to interview whole population. Apart from core questions as in the short form, sample survey questionnaire (long form) with sample size about 15 percent will cover the questions on marital status, qualification, employment, fertility, death, housing.

This approach is used in many countries in the world, but it just applied in the 1989 and 1999 census in Vietnam. But there were only two questions on

fertility and deaths included in the sample with sample size of 5% in the 1989 and 3 % in the 1999.

ii). Based on selection of appropriate technology of data processing and fully utilization of the GSO's informatics facilities, to speed up in rate of provision of census data.

The decentralized method of data processing will be applied. Moreover, GSO consider that the Intelligent Character Recognition (ICR) technology will be used for data capture. The use of ICR might requires a huge amount for the purchasing the facilities such as appropriate scanners, servers, client-PC, software.

Recognizing the difficulty which Vietnam is facing, United National Population Fund (UNFPA) continue to provide assistance to the 2009 Vietnam census taking under the framework of the project. One component of this project is technical support for the conduct of the 2009 population census to collect adequate and precise data on age, gender, and other relevant indicators such as ethnic minorities, migrants, male/female ratio at birth, etc. Support will also include piloting new methods for data collection, processing and analysis.