

CHAPTER 1

INTRODUCTION AND OVERVIEW

Introduction

Census of India is one of the largest operations for collecting data on various items on census basis. It involves considerable efforts right from the development of the questionnaire and instructions, recruitment and training thousands of enumeration staff, actual enumeration work, and compilation and tabulation of data to publishing the report. The census provides a mine of valuable data on housing, population and their characteristics, which are not otherwise available at lower geographic levels. Census organization makes all efforts to make available the tabulations required by various Governmental and non-governmental data users by providing data in a set of standard tables. It is obvious that it can not satisfy the requirements of all users. There would be some cross tabulations not available in the standard set of tables that are required for specific planning/monitoring purpose or for research. There have been regular demands for making available the census data at the individual record level, subject to satisfying the statutory confidentiality clauses, to users to prepare their own cross tabulations. Some efforts were made in the past to provide access to record level data to users. The technological constraints and delay in processing census data resulted in only a few users availing this facility.

Providing Public Use Samples is one of the ways in which use of census data can be increased. With this view, it has been decided to make available to users on a sample basis, data records relating to each household collected during the house-listing of 2001 Census. The confidentiality of individual level information is maintained by excluding all data items leading to individual identity. The Office of the Registrar General, India (ORGI) is making available public use sample data from the Census of India 2001 Houselist records, "*Census of India Housing Microdata Sample (CIHMS)*", as part of the data dissemination program. Documentation of the sampling procedure and estimation methodology is provided here to the users to enable them to properly understand and analyze the sample data.

Two types of CDs containing data are being made available for CIHMS- one at national level (one per cent sample) and the other at State level (five per cent or ten per cent sample). The basic difference between these two sets is that the State sample would allow the user to generate estimates at national, state and district levels, while, the one per cent sample would allow tabulation only at national and state levels. It may be mentioned here that while selecting the 5 per cent sample, it was felt necessary to provide larger sample for small districts of the country. Hence a 10 per cent sample has been selected from such districts for inclusion in the state sample. Only records relating to households are part of the file, i.e., the non residential houses are not included in the data files.

About Census of India Housing Microdata Sample (CIHMS)

The *Census of India Housing Microdata Sample (CIHMS)* has been prepared by the ORGI based on the data collected in the Houselisting Operations in the year 2000 throughout India. The CIHMS files are the text (ASCII) files, containing records of individual households for national or the state sample. By pooling the district level samples drawn independently, the state level sample has been prepared. In a state sample, there may be some districts with a 5-percent sample and some with a 10-percent sample depending upon the size of the districts. The record for each household contains a field showing the weight of the record that would also identify whether it is a part of the 5 per cent or 10 per cent sample. These weights for each household, when applied to the individual records, expand the sample to the relevant total.

The procedure of preparing the state level sample is as follows. A five per cent sampling rate at the district level has been adopted for all districts with a minimum population of 100,000, which covers 571 out of the 593 districts. For the remaining 22 districts with population less than 100,000, which are mainly in the smaller States and Union territories, a higher sampling rate of 10 per cent has been adopted. To maintain the confidentiality of the information about the individual household, the geographic information at the district and higher level only are provided in the state sample while in the national one per cent sample only state identity has been provided. The information on population in the household has also been removed.

To improve the sampling efficiency, the household records within each district were first stratified into homogeneous groups on the basis of place of enumeration, tenancy, type of house, electricity and improved source of drinking water facility as stratification variables. In all, 72 strata are formed in each district. The household records were further sorted by all the geographic codes: Tahsil/ Taluk/ P.S./ Development Block/ Circle/ Mandal, Town/ Village, Ward (for urban), Houselisting Block Number, and Household Number within each of the 72 strata. This sorted frame provides a high level of implicit stratification. Finally, a systematic random sample from this sorted frame has been drawn within each district with a sampling interval of 20 or 10 as the case may be.

A detailed description of variables included in the sampled households and recodes of CIHMS is an important input for the data users which may be seen in Annex 2 of Chapter 2.

This CD also contains software 'CSPRO' for quick tabulations (for installation of the software refer Chapter 3). This software is also available for downloading at the website of the US Bureau of Census.

For ordering and pricing information relating to other similar products, the online catalog at the ORGI Internet site www.censusindia.net can be accessed or the Office of the Registrar General, India may be contacted.