TECHNICAL REPORT

PREPARED BY

THE COMMITTEE CONSTITUTED

FOR

1% AND 5% MICRODATA

Office of the Registrar General, India

Ministry of Home Affairs

2A Man Singh Road

New Delhi

Preface

In order to release 1% and 5% Micro data on Population Enumeration from 2001 Census and Houselisting data from 2011 Census, a committee was constituted vide letter No.F.14011/25/2012 – DDU to prepare the sampling methodology and list of sample keeping in view the reliability of the estimate. The committee comprise of Shri MS Thapa, Additional Director EDP, Shri A K Srivastava, Joint Director and Dr. D K Dey, joint Director, Demography Division. The Committee seat twice and gave their opinion.

Dr. Dey prepared the section relating to Sample size, sample design and related standard error, whereas Shri MS Thapa prepared the software and computer related topics under the supervision of Shri Deepak Rastogi, Deputy Director General.

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 cannot 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 Population Enumeration from 2001 Census and Houselisting data from 2011 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 2010 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.

According to Provisional Population Totals, we 26 districts with population less than 1,00,000 with average population of 61851 or 12940 households and 614 districts with average population of 19,68,380 or 411795 households. Instead of estimating sample size separately for each of the state, if we take 5 percent sample for each of the state population more than 1,00,000, the average number of 5 percent comes out to be close to 20,000 houses/households, which is larger than the size of 16000 at 5 percent level of significance. For states less than 1,00,000 population, average household is 12940, as such, we may go for 10 percent level of significance and the sample size may be 4000 houses/households.

It is therefore proposed to adopt 5 percent sample of houses/households of the state with population more than 1,00,000 satisfying the condition of 5 percent level of error and 4000 houses/households of the state population more than 1,00,000 satisfying the condition of 10 percent level of error.

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, 96 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 96 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 3.

This CD also contains software 'CSPro' for quick tabulations (for installation of the software refer Chapter 4). 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.

Chapter - 2

Concept and Definition

Introduction

It is important for the data users to familiarize themselves with the concepts and the definitions of the terms used for proper evaluation of the data contained in this publication. At the same time, it is all the more important to understand the implications of the terms used at the Census of India 2011 for making meaningful comparisons of the similar data generated by various other agencies within the country and with the data produced by other countries in the world. The concepts and definitions adopted at the **Houselisting and Housing Census** Operations and glossary of terms used for sampling are given separately:

Concepts and definitions adopted at the Houselisting and Housing Census Operations, Census of India, 2011

01. Rural-Urban Areas

The data in tables on Houses, Household Amenities and Assets are presented separately for rural and urban areas. The unit of classification in this regard is 'town' for urban areas and 'village' for ruralareas. In the Census of India 2001, the definition of urban area adopted is as follows:

a. All places with a municipality, corporation, cantonment board or notified town area committee, etc.
b. A place satisfying the following three criteria simultaneously:

i. a minimum population of 5,000;
ii. at least 75 per cent of male working population engaged in non-agricultural pursuits; and
iii. a density of population of at least 400 per sq. km. (1,000 per sq. mile).

For identification of places which would qualify to be classified as 'urban' al villages, which, as per the 1991 Census had a population of 4,000 and above, a population density of 400 persons per sq. km. and having at least 75 per cent o male working population engaged in non-agricultural activity were considered. To work out the proportion of male working population referred to above against b)(ii) the data relating to main workers were taken into account.

Apart from these, the outgrowths (OGs) of cities and towns have also beer treated as urban under 'Urban Agglomerations': Examples of out-growths are railway colonies, university campuses, port areas, military camps, etc. that may have come up near a statutory town or city but within the revenue limits of a villages o villages contiguous to the town or city.

Each such individual area by itself may not satisfy the demographic criteria laid down at (b) above to qualify it to be treated as an independent urban unit but may deserve to be clubbed with the towns as a continuous urban spread. Thus, the town level data, wherever presented, also includes the data for outgrowths of such towns.

02. City

Towns with population of 1,00,000 and above are called cities

03. Building

A 'building' is generally a single structure on the ground. Sometimes it is made up of more than one component unit which are used or likely to be used as dwellings (residences) or establishments such as shops, business houses, offices, factories, workshops, worksheds, schools, places of entertainment, places of worship, godowns,stores, etc. It is also possible that buildings which have component units may be used for a combination of purposes such as shop-cum-residence, workshop-cum-residence, office-cum-residence, etc.

Usually a structure will have four walls and a roof. But in some areas the very nature of construction of houses is such that there may not be any wall. Such is the case of conical structures where entrance is also provided but they may not have any walls. Therefore, such of the conical structures are also treated as separate buildings

04. Census House

A 'census house' is a building or part of a building used or recognized as a separate unit because of having a separate main entrance from the road or common courtyard or staircase, etc. It may be occupied or vacant. It may be used for a residential or non- residential purpose or both.

In certain peculiar situations, the manner in which buildings and census houses were identified for numbering in the field by the enumerators is described hereunder:

Sometimes a series of different buildings are found along a street which are joined with one another by common walls on either side looking like a continuous structure. These different units are practically independent of one another and are likely to have been built at different times and owned by different persons. In such cases, though the whole structure with all the adjoining units apparently appears to be one building, each portion was treated as a separate building and its constituent units as separate census houses.

On the other hand, one may come across cases, particularly in large cities of multi-storeyed ownership flats. In these cases while the structure looks like one building, different persons own the flats.

In case of such multi-storeyed structures, having a number of flats owned by different persons, the entire structure was treated as one building and each flat as a separate census house.

If within a large enclosed area, there are separate buildings owned by differe persons then each such building is treated as a separate building. There can be situation where within an enclosed compound there are separate buildings owned l an undertaking or company or even government that are actually in occupation different persons. For example, Indian Oil Corporation colony where the buildin are owned by the Corporation but these is in occupation of their employees. Ea such building was treated as a separate building. But if in any one of these buildin there were flats in occupation of different households, each such flat was reckon as a separate census house. these were treated as separate census houses. On t other hand, in urban areas, where more than one structure within an enclosed open compound (premises) belonging to the same person, e.g., the main house, t servant's quarter, the garage, etc., only one building number was given for t group and each of the constituent a separate census house number. Only cas where a structure with roof and pillars has come up was treated as a building.

05. Uses of Census Houses

The actual use of census house was ascertained and recorded in full. The different uses of census houses were standardized and grouped into ten categories, as shown hereunder:

i) **Residence:** This category includes houses that were used exclusively for residential purpose.

ii) **Residence-cum-other use:** If the census house was used for residence in combination with one or more other non-residential purpose(s) such as residence-cum- grocery shop, residence-cum-workshop (book binding), residence-cum-boarding house, etc., thesame was indicated together with the detail of non-residential use(s) of the census house.

iii) **Shop/Office:** Census houses exclusively used as shops (where articles are bought / sold on cash or credit) and offices were covered under this category. Instructions were also given to indicate the nature of shop or of the office for which it was being put to use.

iv) **School/College, etc.:**This category included all types of educational institutions and training centers without lodging facilities or any residential use. The nature of school or college or training center was also obtained.

v) **Hotel/lodge /guest house, etc.:**This category covers census houses that were used exclusively for temporary stay or stay in transit and where no person or persons were found living for a period of three months or more. The nature or name of such hotel or lodge or guest house was also provided. If any of such hotel or lodge or guest house was used by the inhabitants as a residence for a period of three months or more then this was considered as an institutional household and covered under the category of houses used as 'residence-cum-other use'. Restaurants and all other eating places not endowed with lodging facilities were covered under the category of 'Shop/Office'.

vi) **Hospital/Dispensary, etc.:** The category includes census houses exclusively used as hospitals, dispensaries, nursing homes and such other health or medical institutes provided they were not eligible to be considered as

institutional households. The nature or names of such institutes were also provided under the appropriate column. If in any of above institutes, the inmates were found living for three months or more, treating the inmates of such an institution as an institutional household, the use of the census house Sometimes it becomes difficult to apply the definition of census house strictly in certain cases. For example, in an urban area, if a flat has five rooms, each room having direct entrance fromthe common staircase or courtyard. By definition, this has to be treated as five census houses. If all these five rooms are occupied by a single household it was not realistic to treat them as five census houses. In such a case, 'singleness' of use of these rooms along with the main house should be considered and the entire flat was treated as one census house. On the other hand, if two independent households occupy these five rooms, the first household living in 3 rooms and the second household occupying 2 rooms, then considering the use, the first three rooms together were treated as one census house and the remaining rooms as another census house. But if each room was occupied by an independent household, then each such room was treated as a separate census house. In case of hostels, hotels, etc., even if the door of each room in which an inmate lives opens to a common verandah, staircase, courtyard or a common room, as it happens almost invariably, the entire hostel / hotel building was treated as one census house. But if such hostels / hotels have out-houses or other structures used for different purposes or the same purpose, then each such structure attached to the main hostel / hotel was treated as a separate census house.

In some parts of the country, in rural areas, the pattern of habitation is such that a group of huts, located in a compound, whether enclosed or unenclosed, is occupied by one household. While the main residence may be located in one hut, other huts may be used for sleeping, as a kitchen, bath room, baithak, etc. Though each of the huts was a separate structure, they form a single housing unit and therefore, have to be treated collectively as one building and one census house. If some of the huts are used by one household and the others by a second household as residence, then the two groups of huts were treated as separate census houses. However, if there were also other huts in the compound used for other purposes and not as part of the household's residence such as, cattle shed, workshed, etc., was categorised as 'residence-cum-other use'

.

vii) **Factory/workshop/workshed, etc.:** If any census house was exclusively used for running a factory or a workshop or used as a workshed, the same was considered under this category. These are places where manufacturing, production, processing, repairing or services takes place or where goods and articles are made and sold. These can be large scale industries / small workshops / worksheds. This category of houses, however, excludes all eating places suchas tea stalls, sweet meat (halwai) shops, restaurants (without lodging facility), dhabhas, etc. which are covered in the category of shop / office.

viii) **Place of worship:**This category includes the places of worship such as temples, gurudwaras, mosques, churches, prayer halls, etc.

ix) **Other non-residential use:**This category covers the census houses used as places of entertainment and community gathering and all other non-residential miscellaneous uses of the census houses not covered under any of the above

categories. They include houses used as cattle-shed, godown, garage, petrol pump, power station, pump house, tube well room, cinema house, museum, stadium, etc. In their case also, the actual use was recorded in the Houselist Schedule.

x) **Vacant:** If a census house was found vacant at the time of houselisting i.e., no person was found living in it and also it was not being used for any other non-residential purpose(s), it was shown as vacant and the reason of vacancy such as ‘under repair’, ‘construction incomplete’, ‘want of tenant’, etc. was also indicated.

06. Condition of Census House

The enquiry relating to condition of census house was made only if the census house was used for residential or partly residential purposes. Again, the condition of census house was ascertained only if a normal household occupied the census house. Since the local conditions vary from place to place and even in rural and urban areas, it was not considered feasible to give any single comprehensive definition for the terms used to determine the condition of census house. The condition of census houses was recorded as ‘good’ or ‘livable’ or ‘dilapidated’ based on the perception and response of the respondent

07. Household

. A 'household' is usually a group of persons who normally live together and take their meals from a common kitchen unless the exigencies of work prevent any of them from doing so. Persons in a household may be related or unrelated or a mix of both. However, if a group of unrelated persons live in a census house but do not take their meals from the common kitchen, then they are not constituent of a common household. Each such person was to be treated as a separate household. The important link in finding out whether itwas a household or not was a common kitchen. There may be one member households, two member households or multi-member households.

08. Institutional Household

A group of unrelated persons who live in an institution and take their meals from a common kitchen is called an Institutional Household. Examples of Institutional Households are boarding houses, messes, hostels, hotels, rescue homes, jails, ashrams, orphanages, etc. To make the definition more clearly perceptible to the enumerators at the Census 2001, it was specifically mentioned that this category of households would cover only those households where a group of unrelated persons live in an institution and share a common kitchen.

09. Head of the Household

The head of household for census purposes is a person who is recognized as such by the household. She or he is generally the person who bears the chief responsibility for managing the affairs of the household and takes decision on behalf of the household. The head of household need not necessarily be the oldest male member or an earning member, but may be a female or a younger member of either sex. In case of an absentee de jure 'Head' who is not eligible to be enumerated in the household, the person on whom the responsibility of managing the affairs of household rests was to be regarded as the head irrespective whether the person is male or female.

10. Scheduled Caste and Scheduled Tribe

Article 341 of the Constitution provides that the President may, with respect to any State or Union territory, specify the castes, races or tribes or parts of or groups within castes, races or tribes which shall for the purposes of the Constitution be deemed to be Scheduled Castes in relation to that State or Union territory. Article 342 similarly provides for specification of tribes or tribal communities or parts of or groups within tribes or tribal communities which are to be deemed for the purposes of the Constitution to be Scheduled Tribes in relation to the various States and Union territories. In pursuance of these provisions, the list of Scheduled Castes and / or Scheduled Tribes are notified for each State and Union territory and are valid only within the jurisdiction of that State or Union territory and not outside.

It was ascertained if the head of the household belonged to a Scheduled caste or a Scheduled Tribe. If in reply to this question, the answer was in the affirmative, the name of caste / tribe to which the head of the household belonged was ascertained. Only if the name of caste / tribe returned by the respondent appeared in the list that was made available to the enumerator, she or he was treated as belonging to Scheduled Caste or Scheduled Tribe. If the head of household returns herself / himself asScheduled Caste, her / his religion was also ascertained. This is mainly because Scheduled Caste can be only from Hindus, Sikhs and Buddhists and not from any other religion. But a member of Scheduled Tribe may belong to any religion. This question was not asked in the case of Institutional Households.

11. Scheduled Caste Household

A household was treated as a Scheduled Caste household when its head belonged to a Scheduled Caste.

12. Scheduled Tribe Household

A household was treated as a Scheduled Tribe household if its head belonged to a Scheduled Tribe.

13. Type of Census Houses

The data on types of Census Houses were not actually collected during the Houselisitng Operations. These have been classified according to the types of material used in the construction of wall and roof of the house entered by enumerator in the Houselist Schedule. The basis of their classification is described hereunder:

Permanent Houses: Houses, the walls and roof of which are made of permanent materials. The material of walls can be any one from the following, namely, galvanized iron sheets or other metal sheets, asbestos sheets, burnt bricks, stones or concrete. Roof may be made of from any one of the following materials, namely, tiles, slate, galvanized iron sheets, metal sheets, asbestos sheets, bricks, stones or concrete.

Temporary Houses: Houses in which both walls and roof are made of materials, which have to be replaced frequently. Walls may be made from any one of the following temporary materials, namely, grass, thatch, bamboo, plastic, polythene, mud, unburnt bricks or wood. Roof may be made from any one of the following temporary materials, namely, grass, thatch, bamboo, wood, mud, plastic or polythene.

Semi-permanent houses: Houses in which either the wall or the roof is made of permanent material and the other is made of temporary material.

Serviceable temporary houses: Temporary houses in which wall is made of mud, unburnt bricks or wood.

Non-serviceable temporary houses: Temporary houses in which wall is made of grass, thatch, bamboo, etc., plastic or polythene.

14. Ownership Status of Census Houses

Owned: If a household is occupying the census house owned by itself and is not making payments in the form of rent to anyone, then the household has been considered as living in owned house. A household living in a flat or a house taken on 'ownership' basis on payment of installments, has also been regarded as owning the house, notwithstanding the fact that all the installments have not been paid.

Rented: A household has been treated as living in rented house if rent is paid or contracted for by the household in cash or even in kind.

Any Other: The households neither living in owned houses nor rented houses have beenincluded in this category. This includes the cases where rent free accommodation is provided to employees by their employers or where the ownership either of the land or of the structure does not belong to the household, i.e., houses constructed on encroached land in unregularised slums or anywhere else.

15. Dwelling Room

A room is treated as a dwelling room if it has walls with a doorway and a roof and should be wide and long enough for a person to sleep in, i.e., it should have a length of not less than 2 metres and a breadth of at least 1.5 metres and a height of 2 metres. A dwelling room would include living room, bedroom, dining room, drawing room, study room, servant's room and other habitable rooms. Kitchen, bathroom, latrine, store room, passageway and verandah which are not normally usable for living are not considered as dwelling rooms. A room, used for multipurpose such as sleeping, sitting, dining, storing, cooking, etc., is regarded as a dwelling room. In a situation where a census house is used as a shop or office, etc. and the household also stays in it then the room is not considered as a dwelling room. But if a garage or servant quarter is used by a servant and if she / he also lives in it as a separate household then this has been considered as a dwelling room available to the servant's household. Tent or conical shaped hut if used for living by any household is also considered as dwelling room.

A dwelling room, which is shared by more than one household, has not been counted for anyof them. If two households have a dwelling room each but in addition also share a common dwelling room, then the common room has not been counted for either of the households.

16. Married Couple

A 'married couple' is formed through marriage. All currently married couples living in the household irrespective of their age have been included in the number of married couples. Only those couples that usually live together in the household even if one of the spouses is temporarily absent are counted. The period of absence in such a case should not exceed three months prior to the date of houselisting.

17. Type and Location of Drinking Water Source

Source type of drinking water: The data were collected on eight types o drinking water source. These are: Tap, Handpump, Tubewell, Well, Tank / Pond Lake, River / Canal, Spring and Any other source. The type of source, which is availed of more during the greater part of the year, has been recorded. Handpump means where the water is taken out manually by operating handpump. Tubewel means where sub-soil water is taken out through electricity or diesel pump.

Within premises: The type of drinking water source is treated within the premises if the type of drinking water source used is located within thehouse o within the premises of the census house where the household lives.

Near the premises: The type of drinking water source is considered 'Nea the premises' only if the type of source used is located within a range of 100 metres from the premises of the census house in urban areas and within a distance of 500 metres in the case of rural areas.

Away: For the households living in urban areas, the type of drinking water source is considered 'Away' if the type of drinking water source used is located beyond 100 metres from the premises of the census house. For households living in rural areas the type of drinking water source is considered 'Away', if they have to cover a distance of more than 500 metres to fetch the drinking water.

18. Source of Lighting

The information on the source of lighting in the houses occupied by the households has been collected on six sources. These are: Electricity, Kerosene, Solar, Other oil, Any other and No lighting. The actual source of lighting used for greater part of the year has been collected for each household.

19. Availability of Latrine within the House

A household was treated as a Scheduled Tribe household if its head belonged to a Scheduled Tribe.

20. Connectivity of Waste Water Outlet

i) **Connectivity to Closed drainage:**The house is considered as having connectivity to closed drainage if it has water outlet connected to a closed drainage to carry away the waste water generated by the household(s) living in it.

ii) **Connectivity to Open drainage:** The house is considered as having connectivity to open drainage if it has water outlet connected to an open drainage to carry away the waste water generated by the household(s) living in it.

iii) **No drainage connectivity:** Such of the houses where waste water outlets are not connected t any drainage system to carry away the waste water are included in this category.

21. Bathroom

The information was collected whether or not bathroom was available to the household within the house. The bathroom was considered as available if the household had an exclusive room used as a bathroom within the house. In cases where more than one household in any census house was sharing a common bathroom, the bathroom was considered as available to all the households living in that census house.

22. Kitchen within the House

Meals could be cooked in a kitchen available as an exclusive and separate enclosure within the house or in a room used for other purposes also or this could be cooked outside the house in the open. If available kitchen was independent of any other use within the house, it was recorded as 'available' and if not as an exclusive room for cooking, then 'not available' was entered. If the available kitchen was shared by two or more than two households living in a common census house, this was considered as available to all the households in the house. If cooking was done in an unenclosed space within the house or its premises, the space used for cooking was not treated as a kitchen and in this case 'cooking in open' was entered. In case due to exigencies of work or any other reason the household was not cooking in the census house, 'no cooking' was recorded.

23. Fuel used for Cooking

The type of fuel used for cooking by the household was ascertained and the one used mostly, was duly recorded. If the household was using more than one fuel for cooking, the predominant fuel used for cooking was recorded. In case, if no cooking was done in the

household, the answer was recorded accordingly. The data are available for the following types of fuels:

1. Firewood; 2. Crop residue; 3. Cowdung cake; 4. Coal / Lignite / Charcoal; 5. Kerosene; 6. LPG; 7. Electricity; 8. Bio-gas; and 9. Any other.

Firewood includes woods of different types such as timber, branches and trunk of trees and shrubs, etc.

Typical examples of crop residue are crop produce such as cotton or pulse stems, paddy or wheat straw, etc.

24. Availability of Assets

The availability of following assets in the household was ascertained and the particular item, if available in the household, 'Yes' was entered and if not available, 'No' was recorded:

1. Radio / Transistor
2. Television
3. Telephone
4. Bicycle
5. Scooter / Motor Cycle / Moped
6. Car / Jeep / Van

The enquiry was limited to their availability and not their number. The above assets were treated as available only if these were in reasonable working order.

Regarding availability of telephone, the concern of the enquiry was limited to its availability and not the source. Therefore, if the telephone was provided to any member of the user household by the government, corporate office, company, etc., this was treated as available. This facility was also considered available if the household was using a cellular mobile telephone.

A car or a jeep or a van was treated as available to the household if this was put to use by the members of the household for their own use.

25. Availing Banking Services

The household was considered availing banking services if its head and / or any other member in the household was availing banking services provided by the bank or post office bank as a holder of any type of bank account. This covers all types of commercial banks such as nationalised banks, private banks, foreign banks and the co-operative banks. The credit and thrift societies did not form part of the banking system and as such these services were not covered under banking services.

Glossary of terms used for Sampling

Accuracy: Quality of survey result as measured by the closeness of the survey estimate to the exact or true value being estimated, estimated by the mean square error. The accuracy is affected by both sampling error and bias.

Allocation of sample: The method used in determining how the sample should be distributed. In stratified sampling, it usually refers to the determination of the number of units selected from each stratum. In cluster sampling, it refers to the decision as to the number of clusters to be selected and the size of sample in each cluster.

Census: Data collection program through which attempts are made to collect information about every element (person, household, farm etc.) in the population.

Characteristic: A variable having different possible values for different individual units of Sampling or Analysis. In a sample survey, we observe or measure the values of one or more characteristic for the units in the sample. For example, we observe(or ask about) the area of land in rice, or the number of cattle on a farm.

Cluster sample: A system of sampling in which the units of analysis of the population are considered as grouped in to clusters and a sample of clusters is selected. The selected clusters then determine the units to be included in the sample. The sample may include all units in the selected clusters or a sub-sample of units in each selected cluster.

Cluster: Small group into which a population is divided to facilitate the data collection. The groups generally are defined so as to help break a large survey area into a workload-sized chunks and/or to reduce travel and administrative costs. Ideally, the units in a cluster should be as heterogeneous as possible.

Coefficient of variation: The relative standard error; that is, the standard error as a proportion of the magnitude of the estimate.

Confidence interval: A range above and below the estimated value, which may be expected to, enclosed the true value with a known probability, assuming no bias.

Design effect: The design effect defined as the ratio of the variance of an estimate based on the actual sample design to the corresponding variance from a simple random sample of the same size and is a measure of the relative efficiency of the design.

Efficiency: A comparative measures of one sample design relative to another with respect to amount of precision produced per unit of cost for a given sample size.

Estimates: A numerical quantity calculated from sample data and intended to provide information about an unknown population value.

Non-Sampling Error: The errors mainly arising at the stages of ascertainment and processing of data is termed as non-sampling error. This type of error is common both to complete enumeration and sample survey.

Population: Any clearly defined set of units (or elements) for which estimates are to be made. The elements can be persons, farms, households, blocks, counties, businesses and so on. Most of our discussion deals with sampling from a finite population, containing a finite number of elements.

Precision: Difference between the sample estimates and a complete count value collected under the sample conditions. This is measured by the sampling error or relative sampling error.

Primary Sampling Unit (PSU): The units making up the sampling frame for the first stage of a multistage sample.

Probability of selection: The chance each unit has of being included in the sample.

Proportionate stratified sampling: A system of selecting a stratified sample in which the same probability of selection is used in each stratum.

Sample: A subset of a population. As used in these chapters, it always refers to a probability sample; i.e., a sample in which each element in the population has a known probability of selection.

Sample design: The sampling plan and estimation procedures.

Sample Survey: A data collection program through which information is collected from a probability selected subset of the population.

Sampling Distribution: The distribution of values of a statistic calculated from all possible samples of the same size from the same population.

Sampling Error: The error arising due to drawing inferences about the population on the basis of observations on a part (sample) of it is termed as sampling error. This error is inherent and unavoidable in any and every sampling scheme.

Sampling frame: The totality of sampling units from which a sample is to be selected. The frame may be a listing of persons of housing units; a file of records; a subdivided map.

Sampling unit: The units to be selected. These may or may not be the same as the same as the units of analysis. For example, to obtain information on persons, one might use a complete listing in a census or a register and select a sample person directly. However, one could also select a sample of households and include in the survey all persons in the selected households. Similarly, one could select complete buildings and include all persons. The choice of the most efficient sampling unit is an important consideration in the design of a survey.

Self-Weighting Sample: In a self-weighting sample, every individual observation has the same probability of selection, and consequently, the same weight.

Simple Random Sampling (also called unrestricted random sampling): This is the simplest type of sampling system. Simple Random Sampling is the procedure of

drawing samples from the population in which the sample is drawn one by one in such a way that the probability of selection of the units from the population in the sample is same at each time. For a sample of size n, each of the possible combinations of n elementary units that may be formed from the population of N units has the same chance of selection as every other combination of n units. Moreover, every element will have the same chance of selection as every other element.

Standard Deviation: The standard error of a simple random sample of size one.

Standard Error: A measure of the extent to which estimates from various samples differ from their expected value. With a reasonable large sample, the distribution of sample results for all possible samples is approximately the normal distribution, and probability statements can be made about how close the sample can be expected to the expected value- the probabilities being expressed in the standard error. The standard error usually is expressed by the Greek letter σ or S.

Statistic: A quantity computed from sample observations of a observations of a characteristic, usually for the purpose of making an inference about the population. The characteristic may be any variable associated with a member of the population; such as age, income, employment status etc. The quantity may be total, an average, a median or other percentile; it may also be a rate of change, a percentage, a standard deviation or any other quantity whose value we wish to estimate for the population.

Stratification: The process of dividing a population into groups for the purpose of selecting a separate sample from each group. Each group is usually made as internally homogeneous as possible. The groups are called strata, with each one referred to as a stratum.

Stratified Sampling: The method of sampling from a universe which has been stratified. At least one sample unit must be selected from each stratum. Probabilities of selection can be different from stratum to stratum.

Stratifying Factor: The criterion, which enables us to classify various sampling units into different strata.

Stratification variables: Variables used for stratification.

Systematic Sampling: A sampling technique in which only the first unit is selected with the help of random numbers and the rest get selected automatically according to some pre-designed pattern. In this method of sample selection, the population is listed in some order and every k^{th} element is selected for the sample.

Non-Sampling Error: The errors mainly arising at the stages of ascertainment and processing of data is termed as non-sampling error. This type of error is common to complete enumeration and sample survey. In practice, the census survey results may suffer from non-sampling errors although these may be free from sampling error. This error can occur at any one or more of the stages of a survey, i.e., planning, field work and tabulation of survey data.

Variance: The square of the standard error, usually, written as S^2 with a subscript to indicate the statistic to which it refers. The term is also used without a subscript for the square of the standard deviation. Where there is any possibility of confusion, sampling variance is used for the square of the standard error, and population variance for the square of the standard deviation.

Chapter - 3
Technical Description

Introduction

This chapter provides the sampling design and sample size for the desired design and the guidelines for the users of the *Census of India Housing Micro-data Sample (CIHMS)* files. It also provides the formulae used for deriving the estimates based on sample data.

Sample Design

The sample design for selecting the CIHMS is based on the objectives of providing sample data to the users for generating reliable tailor made cross-tabulation at affordable cost. This sample would enable the data users to generate dependable estimates for different geographic domains down to the district level. To maintain the confidentiality of the Census Houselist data, the identification particulars below the district level have been excluded from CIHMS files.

The CIHMS comprised of houses which are used as “residence” and “residence-cum- other”, as detailed information on housing assets and amenities are available only for these two categories. The other types of buildings, i.e., shop, school, hotel, factory, etc. are not covered in the CIHMS. The ultimate sampling unit is the household, and the data users can analyze the characteristics at the household level. There are two types of CIHMS files, one for state level sample and the other for national level sample. However, the CD contains only one of the data sets.

Sample Size for State CIHMS

In order to estimate the sample size to be given to the user so the that they can estimate the population parameter at desired level of significance, an exercise has been undertaken to estimate the sample size based on 1) decisive indicator, 2) level of aggregation and 3) the desired level of precision say at 5 percent or 10 percent level of significance. With the objective of estimating the availability of improved water having electricity with tenure status having the pucca/kutcha/semi pucca in rural as well as urban areas, the sample size at 5 percent level of significance has been estimated taking the rates available at all India. The sample size required at 5 percent level of significance with the objective of estimating availability of improved water, availability of electricity for owned pucca houses is 5211 census houses, for kutcha owned houses is 17441 and that of semi-pucca is 8751. For the urban arteas, the values are 681, 28091 and 7416 respectively. At 10 percent level, for rural, these values are 1303, 4360 and 2188 and for urban these are 170, 7023 and 1854. It is seen that the sample size requirement in case of kutcha houses is much larger since the proportion of kutcha houses is just 13.9 percent in rural and 3.1 in urban areas. If we ignore this we need about 16,000 houses/households for 5 percent level of significance or about 4000 houses/households for 10 percent level of significance. Sample size estimation sheet is annexed.

According to Provisional Population Totals, we 26 districts with population less than 1,00,000 with average population of 61851 or 12940 households and 614 districts with average population of 19,68,380 or 411795 households. Instead of estimating

sample size separately for each of the state, if we take 5 percent sample for each of the state population more than 1,00,000, the average number of 5 percent comes out to be close to 20,000 houses/households, which is larger than the size of 16000 at 5 percent level of significance. For states less than 1,00,000 population, average household is 12940, as such, we may go for 10 percent level of significance and the sample size may be 4000 houses/households.

It is therefore proposed to adopt 5 percent sample of houses/households of the state with population more than 1,00,000 satisfying the condition of 5 percent level of error and 4000 houses/households of the state population more than 1,00,000 satisfying the condition of 10 percent level of error.

Sampling Design for State CIHMS

The sample design adopted is a uni-stage stratified systematic random sample. To improve the sampling efficiency, the household within each district were first stratified into homogeneous groups using following variables for the stratification:

(1) Place of enumeration, 2 categories:
1 - Urban
2 - Rural

(2) Tenancy, 3 categories:
1 - Owned
2 - Rented
3 – Other

(3) Type of House, 4 categories:
1 - Pucca (permanent)
2 - Kutcha (temporary)
3 - Semi-pucca (semi-permanent)
4 – Other type houses

(4) Electricity, 2 categories:
1 - Available
2 - Not available

(5) Improved drinking water, 2 categories:
1 - Improved source of drinking water
2 - Other
(The description of the stratification variables is presented **in Annex-3**.)

Using the above stratification variables, 96 strata are formed in each district.

Within each of the 96 strata, the households were further sorted by the geographic codes: Tahsil/ Taluk/ P.S./ Development Block/ Circle/ Mandal, Town/ Village, Ward (for urban), Houselisting Block Number, and Household Number. To ensure a proportional and representative sample from each of the 96 strata within each district, random systematic sampling based on the sorted frame of households was

used. This provides implicit stratification by place of enumeration, tenancy, type of house, electricity, improved drinking water and geography. In the case of a systematic 5 percent sample within a district, a random number between 1 and 20 was selected to identify the first household record to be selected. Beginning with this random start, every 20th household record was selected systematically throughout the entire ordered frame. For the systematic sample of 10 percent in the districts with a population of less than 100,000, a random number between 1 and 10 was selected, and every 10th household was selected systematically. The sample households within a district of a particular state are aggregated to form state level sample. Thus, in a state sample there may be some districts with a 5- percent sample and some with a 10-percent sample.

Sampling Design for National CIHMS

The National CIHMS is an aggregate of 20 percent or 10 percent sub samples of 5 percent sample or 10 percent sample respectively selected at district level for state sample using the same implicit stratification, which effectively give the 1 percent sample of total households at national level. For example, in the case of the districts with a 5 percent sample of households, a random systematic sub sample of 1 in 5 households was selected from the 5 percent sample, keeping the original order of selection based on the stratification criteria. A random systematic sub sample of 1 in 10 households was selected from the 10 percent sample for the smaller districts. The national CIHMS will enable us to generate the reliable estimates at state level only. For this purpose, the identity of the districts has been concealed by replacing their location codes by 99 in the 1 percent national level CIHMS file.

Weighting the Sample Data

As the CIHMS is a sample data from the Census of India 2001 Houselist records, it is necessary to apply a weight or expansion factor to generate the estimates for making inferences for the entire population at districts/state/national levels. The selection of the weight would depend upon the level of aggregation. The weight is calculated as the inverse of the sampling fraction. Therefore, the data records for districts with a 5 percent sampling rate are assigned a weight of 20, and those in districts with a 10 percent sampling rate are assigned a weight of 10. In the case of the 1 percent sample for the National CIHMS file, the weight is 100. The weights are included in the data file for use in the Census and Survey Processing (CSPro) application or with any other software for generation of estimates of population parameter. The weighted estimates of the total number of households with a particular characteristic will be the sum of the estimates for the households with that characteristic. In the case of ratios and proportions, both the numerator and denominator would be estimated separately using corresponding weights. Software such as CSPro and SPSS can automatically weight the survey data once the weight is specified in the application. The CIHMS file contains individual weights for each household, which when applied to the individual records, expands the sample to the relevant total.

Estimation Procedure

Most of the estimates that users will make from the CIHMS will be the weighted total number of households with particular characteristics. Since the total number of households in the frame for the district or state is known, the weighted estimate of the total number of households with a particular characteristic would be equivalent to:

$$\hat{Y} = N \times p,$$

where: $\hat{Y}$ = weighted total number of households in the geographic domain (for example, district or state) with a particular characteristic as estimated from CIHMS

N = total number of households in the frame for the geographic domain, from the Census Houselist 100 percent data

$\hat{p}$ = proportion of households in geographic domain with a particular characteristic as estimated from CIHMS.

The weighted estimates produced from the CIHMS files are subject to two types of errors: sampling errors and non-sampling errors. The sampling errors are due to the fact that the estimates are based on a sample of the Census of India 2001 Houselist data, not on the 100 percent data used to produce the published results. Another sample would give slightly different results. The sampling error is a measure of the average variability among the estimates from all possible samples, so it is an indicator of the precision of the sample estimate. The non-sampling errors include all other types of error, which affect the Census Houselist data, including response error, coding and processing errors. The quality control procedures in the census operations were designed to reduce the non-sampling errors as much as possible.

The sampling error of an estimate is measured statistically by the standard error, or square root of the variance of the estimate. The standard error decreases according the square root of the sample size, so it depends on the size of the population and the sampling rate.

For estimates of weighted totals, the formula for the standard error based on simple random sampling without replacement would be the following:

$$se(\hat{Y}) = \sqrt{N^2 x(1-f)x\frac{\hat{p}x(1-\hat{p})}{n}}$$

This can be simplified as,

$$se\hat{(Y)}_{SRSWOR} = \sqrt{(1-f)x\hat{Y}x\frac{(N-\hat{Y})}{Nf}}$$

$$f = \frac{n}{N}$$

= sampling fraction used for selecting the particular CIHMS (for example, 0.05 or 0.10 for district-level samples)

n = total number of sample households in the CIHMS file in the domain of estimation

N = total number of households in the strata

$\hat{Y}$ = Estimated total

The term (1-*f*) is referred to as the finite population correction factor. It reduces the standard error when the sampling rate is increased. In the case of the 1 percent national sample, it will not make much difference in the calculation of the standard error.

The sampling procedures used for selecting the sample household records for the CIHMS file is actually slightly more efficient than a simple random sample. The stratification in the sample design results in design effects, which are generally less than one. The design effect defined as the ratio of the variance of an estimate based on the actual sample design to by the corresponding variance from a simple random sample of the same size and is a measure of the relative efficiency of the design. Therefore, the standard errors calculated using the formula above can be considered to be slight overestimate of the actual standard errors, but it provides a good approximation for obtaining the standard error. The formulas above were used to produce a generalized table of standard errors, which can be used to obtain the approximate standard error for a particular estimate, given in Anenex-1, from the CIHMS.

While estimating ratios where both the numerator and denominator would be estimated separately using corresponding weights, then the calculation of standard error includes variance and covariance of both the variables. Such variance of the ratio can be calculated using Standard Statistical software packages, like CENVAR.

For the convenience of the users, three tables of generalized standard errors are presented in Annex 1 for estimates based on a 5 percent sample of the Census Houselist records. These standard errors were calculated using the formulas for a simple random sample without replacement.

Table 1 in Annex 1 shows the standard errors for CIHMS estimates of totals. In order to use this table, it is necessary to know the estimate of the total for a particular characteristic and the approximate total number of households in the Census houselist frame for that domain.

Example 1: If the estimated number of households with a roof of grass/thatch/ bamboo and having electricity in a district is 5000 and the number of electrified households in the district is 1,00,000, then the standard error of the estimated total will be calculated from Table 1. First look for the estimated total 5000 in the row and then look for stratum size, the total number of the electrified households, 1,00,000 in

the column. The figure shown at the intersection of this row and the column is 300, which is the approximate standard error of this estimate. We can calculate the 95 percent confidence interval for this estimate as follows:

$$\hat{Y} \pm 1.96 x se(\hat{Y})_{SRSWOR} = 5{,}000 \pm 1.96 x 300 = 5{,}000 \pm 588$$

Therefore the lower limit of the 95 percent confidence interval is 4,412 and the upper limit is 5,588. This means that there is a 95 percent chance that the true number of households with a roof of grass/thatch/bamboo in this district is between 4,412 and 5,588.

Example 2: Suppose one wants to get the state level estimate for the population total, then the state estimate will be the sum of district estimates. The standard error of such an estimate will be the sum of district level estimates is the sum of district estimates. The standard error will be given by the square root of the sum of the variances of the district estimates. If the y_1, y_2, - - -, y_n are the estimated totals at district level for 1st , 2nd , - - -, and nth district respectively and y is the estimated total at state level, then the formula for the standard error of y will be

$$\text{S.E. } (y) = \text{SQRT } (V(y_1)+V(y_2)+.....+V(y_n))$$

where $V(y_1)$, $V(y_2)$,....., $V(y_n)$ are estimated standard error of estimated totals y_1, y_2,.... .y_n at district level.

Suppose there are '3' districts in a region for which we estimate the number of households living in household with grass/thatch/bamboo roof and having electricity we proceed as follows:

Using Table 1, let us create the following table:

Table

S.No. of District	Estimated number of thatched households with electricity (Y)		Total number of electrified households (N)		Estimated Standard Error	
	Rural	Urban	Rural	Urban	Rural	Urban
1	500	250	25000	5000	96	67
2	500	1000	50000	10000	97	131
3	5000	2500	25000	5000	276	154
Total	6000	3750	100000	20000		

The estimated standard error of the total number of households with thatched roof and electricity is 9750. The standard error of the estimated total is

$$\text{S.E.}(9750) = \text{SQRT } (96^2 + 97^2 + 276^2 + 67^2 + 131^2 + 154^2) = 352$$

However, using the simple formula used in the technical note for preparing Table 1, the standard error will be

SQRT (9750 * (120000 - 9750)*(1 - 0.05) / (120000*0.05)) = 413

The difference is mainly due to the effect of the deep stratification adopted on the precision of the estimate. Thus, the standard errors calculated using the formula used in the Table 1 is slight overestimates of the actual standard errors, but it provides a good approximation for obtaining the standard error.

When an estimate or the total number of households in the population is in between the row and column categories in Table 1, the values in this table can be interpolated to obtain an approximate estimate of the standard error.

Example 3: Suppose the sample estimate of the total number of households with a roof of grass/thatch/bamboo for a particular state is 80,000, and the total number of households in the state is 2,000,000. In this case, first one has to interpolate the standard error for estimated total of 80,000 (which lies between the estimated totals of 75,000 and 100,000) in the row categories corresponding to the total number of households in the state 1,000,000 and 5,000,000 (as the 2,000,000 lies in between 1,000,000 and 5,000,000) respectively. Thereafter, the one has to interpolate the standard error for estimated total of 80,000 corresponding to the total number of households in the state 2,000,000 in the column categories. This can be illustrated as follows:

For estimated total of 75,000 and size of population 1,000,000 the corresponding standard error is 1148, whereas, for estimated total of 100,000 and size of population 1,000,000 the corresponding standard error is 1308. Since for the difference of 25,000 in estimated total for the size of population 1,000,000 the corresponding difference in standard error is 160. Therefore, for the difference of 5,000 in estimated total for the size of population 1,000,000 the corresponding difference in standard error would be 32. Thus, the standard error for estimated total of 80,000 corresponding to the total number of households in the state 1,000,000 in the column categories would be 1180. Similarly, one can interpolate the standard error for estimated total of 80,000 corresponding to the total number of households in the state 5,000,000 in the column categories, which would be 1221.

Now, the user has to interpolate the standard error for estimated total of 80,000 corresponding to the total number of households in the state 2,000,000 (which lies in between 1,000,000 and 5,000,000) in the column categories. Since the estimated total of 80,000 and size of population 1,000,000 the corresponding interpolated standard error is 1180, whereas, for estimated total of 80,000 and size of population 5,000,000 the corresponding interpolated standard error is 1221. It means that for the difference of 4,000,000 in the size of population for the estimated total of 80,000 the corresponding difference in standard error is 41 (i.e., 1221-1180). Therefore, for the difference of 4,000,000 in the size of population for the estimated total of 80,000 the corresponding difference in standard error would be 10.25 (i.e, 41*1,000,000/4,000,000). Thus, the standard error for estimated total of 80,000 corresponding to the total number of households in the state 2,000,000 in the column categories would be 1190.25 (i.e., 1180+10.25) which is approximately 1190. As the standard error does not vary linearly according to the sample size, this interpolation gives only an approximated estimate.

Now, we can calculate the 95 percent confidence interval for this estimate as follows:

$$\hat{Y} \pm 1.96 x se(\hat{Y})_{SRSWOR} = 85{,}000 \pm 1.96 x 1{,}180 = 85{,}000 \pm 2{,}312.8$$

Therefore the lower limit of the 95 percent confidence interval is 77,687 and the upper limit is 82,313. This means that there is a 95 percent chance that the true number of households with a roof of grass/thatch/bamboo in this state is between 77,687 and 82,313.

Some users may prefer to calculate the coefficient of variation (CV) or relative standard error of the estimate of a total, which is defined as the standard error divided by the value of the estimate. Table 2 in Annex 1 shows the CVs for estimates of totals from the 5 percent sample data. This generalized table of CVs would be used in a similar way as Table 1, with the value of the sample estimate appearing in the row categories and the total number of households in the geographic domain indicated in the column categories.

The percentage relative standard error (PRSE) of estimated totals has been calculated using the formula:

$$PRSE(\hat{Y})_{SRSWOR} = \sqrt{(1-f) * \frac{(N-\hat{Y})}{\hat{Y} * Nf}}$$

Data users may also be interested in obtaining sample estimates of the proportion of households with a particular characteristic. In the case of the sample estimate of a proportion, the standard error based on simple random sampling without replacement would be the following:

$$se(\hat{p})_{SRSWOR} = \sqrt{(1-p) * \frac{\hat{p} * (1-\hat{p})}{Nf}}$$

Table 3 in Annex 1 presents the approximate standard errors for estimates of percentages based on a 5 percent sample of Census Houselist records, calculated using the formula above. In the row categories the percentage and its complement are grouped together since they have the same standard error. The column categories refer to the total number of households in the geographic domain, similar to Tables 1 and 2.

The standard error for proportion $p\hat{}$, is the same as its counterpart (1- $p\hat{}$). The difference is, however, in the relative standard error.

The relative standard error (RSE) for $p\hat{}$ is given by the formula

$$RSE(\hat{p})_{SRSWOR} = \sqrt{(1-f)x\frac{(1-\hat{p})}{Nf\bar{p}}}$$

The relative standard error for (1- $p\hat{}$) is given by the formula

$$RSE(1-\hat{p})_{SRSWOR} = \sqrt{(1-f)x\frac{\hat{p}}{Nf(1-\bar{p})}}$$

Data format

The 2011 Census of India Housing Microdata Sample (CIHMS) data files are available in ASCII format (text files). Users of the CD-ROM may access the CIHMS data with any standard statistical software packages. The CD-ROM with software is designed to perform basic cross tabulations of any desired set of variables on the CIHMS file. The CIHMS files are accompanied by a CSPro data dictionary (see Annex- 2) in a format that will allow the user to tabulate data using CSPro Software. The dictionary for India is named "HLINDIA00.dcf", whereas, for the respective states these are named as "HLINDIA01.dcf" (for Jammu & Kashmir), "HLINDIA02.dcf" (for Himachal Pradesh) etc. Separately an additional file, 'INDIA.ANM' giving area names of all States/Union territories and Districts with corresponding codes, as per Census of India 2001, is also made available. Also, these files can be used in getting the desired results for the State and districts with name while doing tabulations through CSPro. In case the user wants to use some other software for data processing, the CSPro data dictionaries are to be transformed accordingly and the corresponding descriptions and values to the categories of each variable required by the software package of choice has to be prepared by the users. For this purpose, the details of every data item are given below:

Data Item provided in CIHMS	Valid Field Value / Range	Starting position in CIHMS	Field Length
State	01-35(refer Annex-5)	**1**	**2**
District	01-70	3	2
Place of enumeration (Rural/Urban)	1 Rural 2 Urban	5	1
Ownership	1 Owned 2 Rented 3 Any Other	6	1
House Type	1 Permanent 2 Temporary 3 Semi-permanent 4 Other types	7	1
Electricity	1 Light available 2 Light not available	8	1
Improved source of Drinking water	1 Improved Source of Drinking Water 2 Other Source of Drinking Water	9	1

Serial Number	0 to 9	10	1
Floor	1 Mud 2 Wood/Bamboo 3 Burnt Brick 4 Stone 5 Cement 6 Mosaic/Floor Tiles 7 Any other material	11	1
Wall	1 Grass/Thatch/Bamboo, etc. 2 Plastic/Polythene 3 Mud/Unburnt brick 4 Wood 5 Stone not packed with Mortar 6 Stone packed with Mortar 7 G.I/Metal/Asbestos Sheets 8 Burnt Brick 9 Concrete 0 Any other material	12	1
Roof	1 Grass/Thatch/Bamboo/Wood/Mud, etc. 2 Plastic/Polythene 3 Hand Made Tiles 4 Machine made Tiles 5 Burnt Brick 6 Stone 7 Slate 8 G.I/Metal/Asbestos Sheets 9 Concrete 0 Any other materials	13	1
Purposes	1 Residence 2 Residence – cum- other use	14	1
Condition	1 Good 2 Livable 3 Dilapidated	15	1
SC/ST	1 SC 2 ST 3 Others	16	1
Dwelling Rooms	0,1,2,.....99	17	2
Married Couples	0,1,2,.....99	19	2
Water Source	1 Tap water from treated source 2 Tap water from untreated source 3 Covered well 4 Un-covered well 5 Hand pump 6 Tubewell/borehole 7 Spring 8 River/canel 9 Tank/pond/lake 0 Other source	21	1
Water Premises	1 Within the premises	22	1

	2 Near the premises 3 Away		
Light Source	1. Electricity 2. Kerosene 3. Solar 4. Other oil 5. Any other 6. No lighting	23	1
Latrine	1. Pile sewer system 2. Septic tank 3. Other system 4. Pit latrine with slab/ventilated improved pit 5. Pit latrine without slab/open pit 6. Night soil disposed into open drain **Service latrine** 7. Night soil removed by human 8. Night soil serviced by animals **No latrine within premises** 9. Public latrine 0. Open	24	1
Waste water	1. Closed Drainage 2. Open Drainage 3. No Drainage	25	1
Bathroom	1. Yes 2. Enclosure without roof 3. No	26	1
Kitchen	1. Cooking inside house has kitchen 2. Cooking inside house does not have kitchen 3. Cooking outside house has kitchen *4.* Cooking out*side house does not have kitchen* 5. No cooking	27	1
Fuel	1. Firewood 2. Crop residue 3. Cowdung cake 4. Coal/lignite/charcoal 5. Kerosene 6. LPG/PNG 7. Electricity 8. Bio-gas 9. Any other 0. No cooking	28	1
Radio/Transistor	1. Yes 2. No	29	1
Television	1. Yes 2. No	30	1
Computer/Laptop	1. Yes with internet	31	1

	2. Without internet 3. No		
Telephone/Mobile phon	1. Yes, Landline only 2. Mobile only 3. Both 4. No	32	1
Bicycle	1. Yes 2. No	33	1
Scooter/Motor Cycle/Moped	1. Yes 2. No	34	1
Car/Jeep/Van	1. Yes 2. No	35	1
Availing Banking Service	1. Yes 2. No	36	1
Weight		37	3
Sampling rate		40	3

File and Record Description

The CIHMS files are of two types, one for the state level and the other for the national level sample. Both the state and national CIHMS contains houselisting records of 42 character length. In the State CIHMS, each record can be identified uniquely down to district level, while, in the national CIHMS, records can be identified down to state level only and the household are the ultimate units at this level.

In the State CIHMS, the identification particulars of households below district level are not provided in the data file. This is a sorted file with selected variables from the actual houselisting data file in each district. The sorting has been done by place of enumeration, tenancy, type of house, electricity, improved drinking water source and geography. The sorting has been done to get an implicit frame for sampling. Each line of record represents the basic information on selected variables of household characteristics of each household selected within the sample. Only residential and residential-cum-other houses are covered. Vacant and institutional houses etc. are excluded from the data file. Each record has been assigned housing unit weight depending on the size of the district.

The national CIHMS will enable us to generate the reliable estimates at state level only. For this purpose, the identity of the districts has been concealed by replacing their location codes by 99 in the 1 percent national level CIHMS file. The same implicit stratification as used for State CIHMS has been maintained for obtaining the national CIHMS.