

Handling Large-Scale Unit Level Data Using STATA
Professor. Pratap C. Mohanty
Department of Humanities and Social Sciences
Indian Institute of Technology, Roorkee
Lecture No. 04
Understanding Unit Level Data: IHDS

Welcome friends once again to the MOOC module on handling large-scale unit level data with STATA. This is our week number 1 where we have been discussing on unit level data. We are trying to familiarize with the existing unit level data, which are largely used in Indian context. So, this module, this particular lecture is targeted to understand India human development survey. In the previous lecture I think I have already explained the use of or the constitution of NSS; where the type of data they provide type of questions, type of layout, type of character spacing, type of some of the variables, the timeline we discuss in NSS.

Similarly, we also unfold the newness in the data and the quality of the data in IHDS. How this is completely or to a large extent different than that of the existing datasets. So, it is going to be very interesting for you, and this I have been using over last couple of years for my research. And also for my PhD students and they are happy in using this IHDS dataset; especially those who are interested in understanding quality of human life. especially, understanding poverty in a different context maybe you can say multi-dimensional context and that too for policy framing.

Most importantly IHDS is interesting to be noted; whereas some direction can be derived some important features can be derived for policy framework. So, your results out of the IHDS is going to be very very interesting and could be referred widely. So, let me have the privilege to unfold the details with the first slide; this is where we start with IHDS. IHDS is a joint collaboration of University of Maryland.

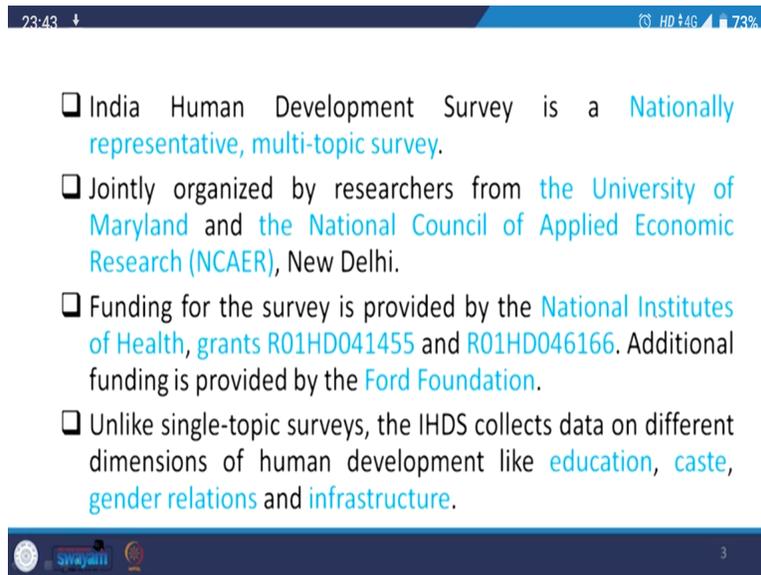
(Refer Slide Time: 02:47)



And NCAER is one of our institute located in Delhi; which is famously known as social science institute. And the team member the professor who is famous in developing these datasets is professor Sonalde Desai, very very articulatively structured the data; I am using articulatively because many of the information are very indicative. And it can be interpreted in various context; so let me go into the deeper understanding of this data.

So far as IHDS is concerned, first starting information for all of you is that this is also a nationally representative dataset; because of its multi-structure coverage, because of its multi-layering weightage, because of its multi issues coverage.

(Refer Slide Time: 03:54)



- ❑ India Human Development Survey is a **Nationally representative, multi-topic survey**.
- ❑ Jointly organized by researchers from **the University of Maryland** and **the National Council of Applied Economic Research (NCAER)**, New Delhi.
- ❑ Funding for the survey is provided by the **National Institutes of Health**, grants **R01HD041455** and **R01HD046166**. Additional funding is provided by the **Ford Foundation**.
- ❑ Unlike single-topic surveys, the IHDS collects data on different dimensions of human development like **education, caste, gender relations** and **infrastructure**.

And as I just discussed National Council of Applied Economic Research institute (NCAER) and University of Maryland is behind the think tank of India human development survey. It is funded by National Institutes of Health (NIH); world usually give top bigger funding with the grant numbers are given here for your reference; additional funding is also provided by Ford foundation. So, unlike single topic surveys the IHDS collects data on different dimensions of human development like education, caste, gender relations and infrastructure. It is not just gender in isolation; it is a gender relationship.

Basically, where we will be emphasizing quality of gender: gender rights, gender decisions, gender empowering gender. So, a gamut of information is available for your work. The foundation of the IHDS lie in work with NCAER. NCAER which had done in 1993-94 as part of its human development profile of India; famously known as HDPI. The NCAER initially started with its survey on this IHDS framework in 1993; so IHDS-1, that has been carried forward to the next round, in another round. That is called IHDS-1 of the rural sample used about a one third of the HDPI.

(Refer Slide Time: 05:30)

- ❑ The foundations of IHDS lie in work NCAER had done in 1993-94 as part of its **Human Development Profile of India (HDPI)**. The IHDS-I rural sample used about a third of the HDPI households.
- ❑ IHDS public data files are available from the Data Sharing and Demographic Research (DSDR) of Inter-university Consortium for Political and Social Research (ICPSR).

Data access:

<https://www.icpsr.umich.edu/icpsrweb/DSDR/search/studies?q=ihds+data>



So, IHDS is famously known from its first round and second round; not in the HDPI round. So, HDPI is initially by the NCAER only and not with the collaboration of University of Maryland. So, IHDS carried the rural sample and that to one third of the information of the rural sample in HDPI. So, that means those who are interested in having a repeated information or longitudinal information related to rural profile. They can start from the work of 1993-94 till the latest round of 2011; that is IHDS-2, the second round of IHDS.

So, in totality for understanding rural space or rural changes over time; you can consider 3 rounds of IHDS, so HDPI is considered and IHDS-1 and IHDS-2. I will discuss in detail in our successive slides. So, IHDS is a public data. IHDS public data files are available from the Data Sharing and Demographic Research (DSDR). Famously known as DSDR of Inter-University Consortium for Political and Social Research; it is famously also known as ICPSR. The website is linked here, you can explore further details and there are also many information for your knowledge.

(Refer Slide Time: 06:54)

- ❑ The only **Longitudinal dataset at unit level** available in India.
- ❑ Its coverage permits analysis of associations across a range of **social and economic conditions**.
- ❑ Two rounds of the survey have been conducted since. ^{HDPI} (IHDS-I - II)
- ❑ IHDS is the **first large-scale national panel survey** of over **40,000** Indian rural and urban households undertaken by researchers from NCAER and the University of Maryland.

As I have already mentioned so far no single survey in India which discussed about longitudinal dataset at unit level. This is the only existing database in India which trust upon the longitudinal characters, repetitions of the same person; some of the percentage of the person will discuss that in the next round of a survey. Since it is longitudinal, all other surveys are usually called time series based; or it is not time series or cross sectional data. Purely cross sectional repeated in over time, not by the particular unit of the person.

But the same person is repeated in other round; but not exactly 100 percent, it is less than that, will discuss in detail. And the coverage of these IHDS permits analysis of associations across a range of social economic conditions. Two rounds of the survey have been conducted since then; that is IHDS, IHDS-I and II. So, if you mix with the rural, then we can include HDPI as well, as I already mentioned. So, IHDS is the first large-scale national level panel survey of over 40,000 Indian rural as well as urban households; undertaken by researchers from NCAER and University of Maryland.

40,000 Indian households you can just simply multiply by roughly around 4 to 5 household members; down these boils down to 2 lakhs plus observations. Which is in fact one of the very good platform to be defined as national representative, so far as this number is concerned.

(Refer Slide Time: 08:50)

WHAT MAKES IHDS UNIQUE AMONG INDIAN SURVEYS

□ Breadth of Topics

- A wide range of topics are covered in a single survey.
- This breadth permits analyses of associations across a range of social and economic conditions.
- Examples are: studying children's outcomes (e.g., learning, immunizations) requires joint consideration of the role of poverty, family structure, gender relations, community context, and the availability of facilities.
- The questionnaires are similar across the two waves to enable comparisons over time. However, there are additions to the institutional modules to capture new programs and policies, which provide a quasi-experimental framework to test for their effectiveness



What makes IHDS unique among Indian surveys? If you wish to understand a bit. So, first important point is the topic covered the breadth of the topic covered; there are wide range of topics covered in this single survey. The breadth permits analysis of associations across a range of social and economic conditions. For example, it covers children's outcome like learning, immunization related information, nutrition related information; so which requires joint consideration of the role of poverty as well as family structure, gender relations, community context and some other facilities available to the households.

Similarly, the questionnaire is discussed which are similar across the two waves, we are using waves; because of the fact that the time is different and the portions are same not by 100 percent. We will discuss the exact number in our next slide. Waves are discussed when we have a panel component or longitudinal component; which enable comparison over time. However, there are additions to the institutional modules to capture new programs and policies, which provide a quasi-experimental framework.

I think it is needless to mention that why the top most research journals; which are having better impact factor or quite refereed in international context consider experimental papers, design based papers, then that of a cross sectional one. They give better weightage to it; because it resulted in a holistic direction, holistic interpretation which could be interpreted for policy. And

recently if you go by the Noble prize in economics; it was given to professor Abhijeet Banerjee for using the module called Randomized Control Trial (RCT).

(Refer Slide Time: 11:06)

□ Depth of Human Development Indicators

- Along with traditional social and economic indicators (e.g. poverty, school enrolment etc.) IHDS is supplemented with **additional measures** that provide a fuller understanding of actual human development.
- IHDS measures not only school enrollments but children's **actual achievements in reading, arithmetic, and writing**.
- the usual consumption based measures of living standards are supplemented with detailed questions on the **levels and sources of income** in order to better understand how the household came to its current economic position.



RCT is useful technique in the form of trial because it is a control format and and some experiments are designed. So, similarly though it is not RCT as such, we are just mentioning in terms of some forms of experiment; and these it is not 100 percent experimental; it is called quasi-experimental. We will clarify in some of the models, what forms these are called quasi-experimental; not exactly experimental to test their effectiveness in other context.

Looking at the depth of the human development indicators along with the traditional socioeconomic indicators like poverty, school enrolment, IHDS is supplemented with additional measures and that provide a full on understanding of the actual human development indicators. It measures not only school enrolment but also children's achievement related to some cognitive issues in terms of some say it is related to reading, arithmetic skill, writing habits are also covered.

The usual consumption based measures of living standards are supplemented with detailed questions which are on the levels and sources of income, in order to better understand how the household came to its current position.

(Refer Slide Time: 12:34)

A Panel Component

- This dataset provides a rich description of changes in the Indian society.
- IHDS-1 is a nationally representative survey of 41,554 households conducted in 2004-05.
- IHDS-2 re-interviewed **83 per cent** of the household as well as split households to trace changes in their lives.
- Panel data will allow better estimation of the influence of development and government policy on families and households.



8

Understanding the panel component a bit; the data set provides a rich area of discussion on changes to the Indian society over time. IHDS-1 is a national representative survey which covers 41,554 households which was conducted in 2004-05. And these are re-interviewed in IHDS-2 in our second round; re-interviewed percentage were of 83 percent of the households. And the second round also considered some split households, in addition to the re-interviewed household. So, because of the fact that there are some family which are splitted two different members; not necessarily the same family carried over time.

So, some new members are also added either, that is also called split, or a new none other than the family members are also considered in the second round of the survey. So, these are called split not necessarily just continue with the 100 percent panel; it is called 83 percent panel. So, if you go by 83 percent panel and if we filter it to 83 percent in our new data set, then that will be called 100 percent panel. Isn't it? So, panel data allow better estimation of the influence of development and government policy I have already discussed.

(Refer Slide Time: 13:56)

□ A Rich Array of Contextual Measures

- In addition to the household surveys, IHDS includes **village**, **school**, and **medical facility** surveys.
- Extensive Census data are also available to be merged for contextual analyses at the village, district, and state levels.
- Within the household survey, several sections also focus on the household's connections to the wider community.



It is a rich array of contextual measures, because of the fact that it includes village information, school information, medical facilities etc. Extensive census data are also available to be merge for the contextual analyses at the village, district, and state level. Within the household survey, several sections focus on the household connection to the wider community; so, in order to make it more representative as well.

What is the sample frame of the IHDS survey? We have already discussed the sample frame in our previous lecture for NSS. let us understand the sampling design part, where sampling frame carries the most important portion.

(Refer Slide Time: 14:45)

SAMPLING FRAME OF IHDS SURVEY

□ Sampling frame of IHDS-1

- Was conducted in all the Indian states and union territories (except **Andaman Nicobar** and **Lakshadweep**)
- Nationally representative survey of **41,554** households consisting 26,734 rural and 14,820 urban households.
- The rural sample was drawn using **stratified random sampling** and contains 13,900 rural households who were interviewed in 1993-94 HDPI survey and 27,654 new households.
- The urban sample was a **stratified sample of towns and cities** within states or group of states selected by **probability proportional to population (PPP)**.



10

It was conducted in all the Indian states and union territories except two that is Andaman Nicobar and Lakshadweep. So, this is a national representative survey covers 41,554 households consisting of 26,734 rural and 14,820 urban households. The rural was taken using a stratified random sampling, which contains over 13,900 rural households; who were interviewed in 1993-94 considered in the next round; that is in the IHDS-2; we are referring to rural who has drawn from certified sampling.

If you are trying to make a panel based on the HDPI of 1993-94; so 13,900 rural households are interviewed. Alright, from the HDPI survey 27, 654 new households are added in IHDS-1. And the urban sample was a stratified sample of towns and cities within states or group of states selected by the technique of sampling called PPP (probability proportional to population). Sometimes it is also called probability proportional to size sampling, size refer to population.

(Refer Slide Time: 16:16)

- ❑ Out of 593 districts in India in 2001, 384 are included in IHDS.
- ❑ The sample is spread across 1503 villages and 971 urban blocks.



Out of the 593 districts in India in 2001; 384 are included for the IHDS survey. The sample is spread over a 1503 villages and 971 urban blocks. So, regarding urban blocks and all we discuss some ideas in NSS; but not necessarily the same is carried forward in IHDS; you have to mark carefully for the IHDS.

(Refer Slide Time: 16:47)

❑ Sampling frame for IHDS-2

- ❑ Total number of household surveyed in IHDS-2 is 42,152 households that consists of 27,579 rural and 14,573 urban.
- ❑ IHDS-2 re-interviewed 83 per cent of original households of IHDS-1 as well as split households residing within the village and an additional sample of 2,134 households.
- ❑ These households are spread across 33 states and union territories, 384 districts, 1420 villages and 1042 urban blocks



- ❑ Out of 593 districts in India in 2001, 384 are included in IHDS.
- ❑ The sample is spread across 1503 villages and 971 urban blocks.

So far as the sampling frame for IHDS-2 is concerned, in the previous phase I discussed IHDS-1, we are discussing IHDS-2. So, in IHDS-2, total number of household surveyed was of 42,152 households; and that consists of 27,579 rural and 14,573 urban. IHDS-2 re-interviewed aspects are discussed in the previous slide, I said for rural summary interviews are there but from the HDPI. But we are saying how many are re-interviewed in total; not just rural and urban that is of 83 percent. 83 percent of the original household those are covered in IHDS-1, is re-interviewed in the IHDS-2.

In addition to that there are also split households residing within the village. An additional sample of 2,134 households. So, these households are spread across 33 states and union territories, 384 districts as I just said were covered. I think here we have 384 mentioned; 300 districts, about 1420 in the first round it was of 1503 villages. It is a little less, 1420 villages and 1442 urban blocks; it is higher than that of the previous. This one is higher as compared to the previous round and so far as village it is lesser.

(Refer Slide Time: 18:30)

Final Sample of Households

| | |
|---|--------|
| Households in IHDS | 41,554 |
| Households in IHDS-II | 42,152 |
| Households surveyed in both IHDS and IHDS-II | 40,018 |
| IHDS households lost to recontact for IHDS-II | 6,911 |
| IHDS-II households not included in IHDS | 2,134 |



13

These are the final sample of the households; so household IHDS-1 41,554 I have already said. Household in IHDS-2 it is of 42,152 and household survey in both the rounds IHDS-1 and 2 was a 40,018; this is very important those who are interested to make panel. IHDS households lost to re-contact for the IHDS-2 are of 6,911; sometimes we mentioned as attrition rate. IHDS-2 households not included in IHDS that is of 2,134.

(Refer Slide Time: 19:23)

SURVEY METHODOLOGY OF IHDS DATA

Face to face interviews were carried out to collect information on different modules:

- An interview with a knowledgeable informant-
 - Typically **head of the household**.
 - Socio economic condition** of the household like income, employment, educational status, consumption expenditure and social capital.
- An interview with an ever-married women-
 - Aged **15-49**
 - Regarding health, education, fertility, family planning, marriage and gender relations in the households and community.
 - Those ever-married women who were interviewed in IHDS-1 but were no longer eligible (older than 49 years) were also interviewed.

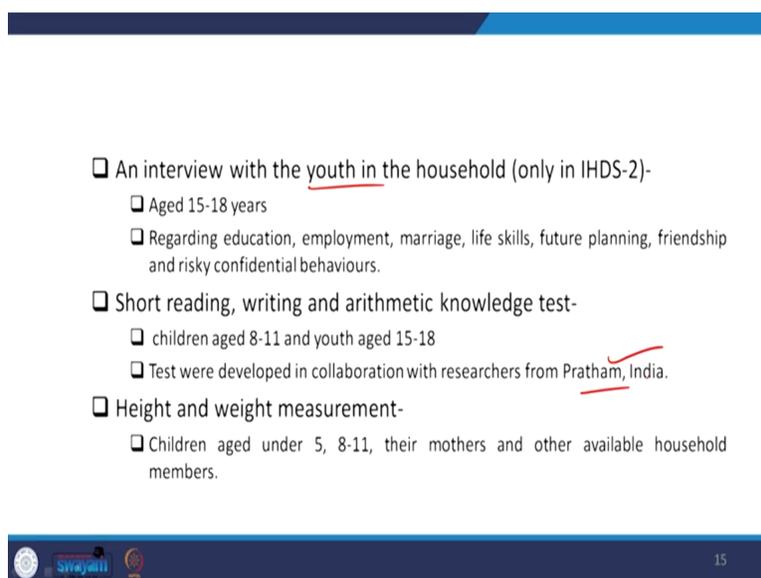


14

So looking at the survey methodology of IHDS data, face to face interviews are conducted to collect information on different modules. An interview with a knowledgeable informant that is typically head of the household covered. Socio-economic condition of the household like income, employment, education, consumption expenditure, social capital; an interview with an ever-married women discussed in the IHDS-2. Though it was clubbed in IHDS-1 with the household file, when we discussed for ever-married women; we refer to the age group of 15 to 49. Those are also called also called eligible women.

In the file the exact name is called eligible women, who go or experience the reproductive age that is of 15 to 49. Regarding health, education, fertility, family planning marriage, gender relations in the households and the community etc. are covered. Those ever-married women who were interviewed in IHDS-1; but were no longer eligible, older than 49 years were also interviewed; even if they are older or also interviewed in the next round.

(Refer Slide Time: 20:55)



- An interview with the youth in the household (only in IHDS-2)-
 - Aged 15-18 years
 - Regarding education, employment, marriage, life skills, future planning, friendship and risky confidential behaviours.
- Short reading, writing and arithmetic knowledge test-
 - children aged 8-11 and youth aged 15-18
 - Test were developed in collaboration with researchers from Pratham, India.
- Height and weight measurement-
 - Children aged under 5, 8-11, their mothers and other available household members.

So, we can make a comparison of their status those who have already crossed their reproductive age or reproductive duration. So, an interview with the youth in the household were only covered in IHDS-2 regarding youth information were not covered in IHDS-1; this is very unique so far as India's dataset is concerned. So, those who are having age 15 to 18 years, these covers this particular file covers education, employment, marriage, life skill, future planning, friendship, risky confidential behaviours.

Similarly, short reading, writing skills, arithmetic knowledge related test, capability or the skill; knowledge related information of the children are also covered. Children with age 8 to 11 and youth age 15 to 18 discussed in this particular regard like for reading, writing, arithmetic knowledge. Test were developed in collaboration researchers from Pratham; Pratham is one of the famous NGO, who collaborated for the support. Height and weight measurement were made for children under 5, 8 to 11, their mothers and other available household members; their details are also covered.

(Refer Slide Time: 22:18)

- Facilities assessment of one government and one private primary school as well as a primary health care facilities in the community.
- Village questionnaire assessing employment opportunities and infrastructure facilities in the village.



Facilities assessment of the one government and one private primary school as well as a primary health care facility in the community are also covered. Village questionnaire assessing employment opportunities and infrastructure facilities in the village is very useful; so far as village facilities and policy facilitations are concerned, if you request the data for this particular one specific data information maybe provided by IHDS to the individual researcher.

(Refer Slide Time: 22:51)

DIFFERENT WAVES OF IHDS DATA

India Human Development Survey-1

- The first round was completed in 2004-05
- IHDS 1 data are assembled in 8 datasets:
 - DS0 Study-level file ✓
 - DS1 Individual ✓
 - DS2 Household ✓
 - DS4 Medical ✓
 - DS5 Non-resident ✓
 - DS6 Primary School ✓
 - DS7 Birth history ✓
 - DS8 Crops ✓
- Weblink:
<https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/22626/datadocumentation>



So, let us consider what files exactly it covered different waves in and their structure by their files availability. The first round was completed in 2004-05 already discussed. It contains 8 different datasets like DS0 study-level file, individual, household, medical, non-resident, primary school, birth history and crops related information also given in the first round first wave. And the detail you can follow the web link given here.

(Refer Slide Time: 23:27)

India Human Development Survey-2

- Complete in 2011-12 ✓
- IHDS-2 data are assembled in 14 datasets:
 - DS0 Study-level file
 - DS1 Individual
 - DS2 Household
 - DS3 Eligible women
 - DS4 Birth history
 - DS5 Medical staff
 - DS6 Medical facilities
 - DS7 Non-resident
 - DS8 School staff
 - DS9 School facilities



- DS10 Wage and salary
- DS11 Tracking
- DS12 Village
- DS13 Village panchayat
- DS14 Village respondent

Weblink:

<https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/36151/data/documentation>



Similarly, for the IHDS-2 there are 14 datasets covered; 14 datasets we have already discussed that some datasets are new in the sense that differently mentioned like eligible women discussed in the first round within the file household. But, here it is eligible similarly youth profile is also very new, so far as we have two is concerned; and this was conducted in 2011 and 12 and completed in two 2011-12. There are 14 files available, so you can follow this and try to understand the details. The respective link is given here.

(Refer Slide Time: 24:08)

WORKING WITH UNIT LEVEL IHDS DATA

To understand the unit level data provided by IHDS following files are needed:

- The questionnaire file**- set of items on which the information is required.
 - For IHDS-1, Household questionnaire, education and health questionnaire and learning test are available in Pdf format.
 - For IHDS-2, Income and social capital questionnaire (Household questionnaire in IHDS-1), education and health questionnaire and learning tests are available in pdf format.



(Refer Slide Time: 26:07)

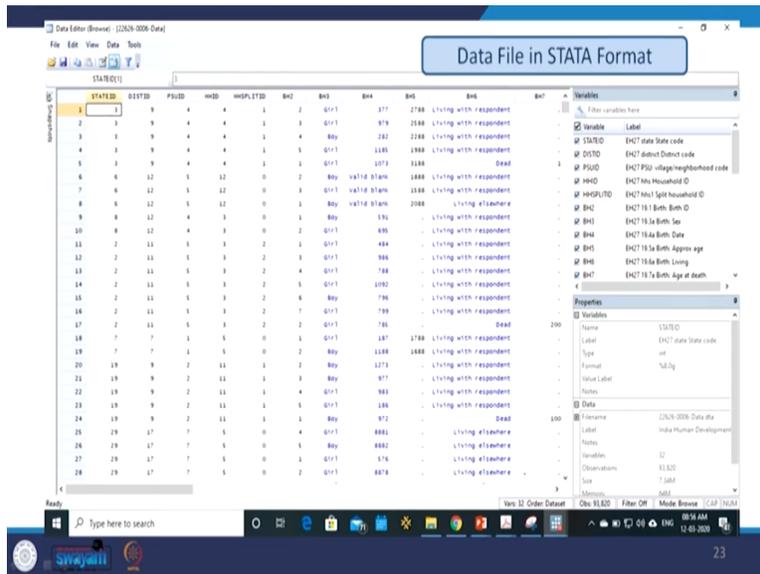
□ The data file- In different formats the Data files are available.

- [Questionnaire \[PDF\] original MULTI](#)
- [Codebook \[PDF\]](#)
- [Stata](#)
- [R](#)
- [SPSS](#)
- [SAS](#)
- [Delimited](#)
- [ASCII](#)
- [ASCII + Stata Setup](#)
- [ASCII + SAS Setup](#)
- [ASCII + SPSS Setup](#)



The data file in different formats are available whoever wants to work with STATA; they can download the data by clicking here on STATA. If you are interested to work with R; you can download the extracted data in R format or SPSS format, SAS format there are different formats. Even the ASCII data which we discussed in the couple of lectures before is also discussed.

(Refer Slide Time: 26:39)



This is the format data file in STATA format, if you download in STATA format; then this is the snapshot in STATA format.

(Refer Slide Time: 26:51)

The codebook file-

- describes the summary of variables like, maximum/minimum value, unweighted frequency.
- Character type of variables.
- Variable range
- Frequency of missing values.



After opening that particular file, the codebook files are given like here; codebook describe the summary of variables. Its maximum/minimum value, unweighted frequency is also given, character type of variables, variable range, frequency of the missing values. How to deal with the variable is very important; it may be the case that there are so many missing values. So, correlating with some other information and handling the missing value is very very important. This is the codebook we are referring to it gives a complete Pdf document regarding the codebook.

(Refer Slide Time: 27:35)

```
. codebook, tab(36)

-----
stateid                               State code
-----

type: numeric (byte)
label: stateid
range: [1, 34]
unique values: 33

units: 1
missing ..: 0/41554

tabulation: Freq.  Numeric  Label
             -----
             715         1  Jammu & Kashmir 01
             1372        2  Himachal Pradesh 02
             1593        3  Punjab 03
             90          4  Chandigarh 04
             458         5  Uttaranchal 05
             1618        6  Haryana 06
             960         7  Delhi 07
             2485        8  Rajasthan 08
             3512        9  Uttar Pradesh 09
             1430       10  Bihar 10
             105        11  Sikkim 11
             105        12  Arunachal Pradesh 12
             130        13  Nagaland 13
             105        14  Manipur 14
             105        15  Mizoram 15
             229        16  Tripura 16
             161        17  Mehalava 17
```

It gives the range information, maximum range of that codes and the missing information is given. Here it gives, there is no value as there is no missing, the label the name of that particular variable is state id its numbers are given, different codes are given and the entries are mentioned. And also some frequency information is also given in the codebook file itself; but we can easily derive from our STATA file. But, just by getting a tab on it; so will get the exact result.

(Refer Slide Time: 28:08)

The user guide-

- Describes how the data is organized.
- Sampling of the survey.
- Merging and linking of files
- Interpretation of missing values.
- Methodology of the survey
- Unique IDs
- Constructed variables, etc.

So, we need not worry this moment. So far as the user guide is concerned; it describes how the data is organized, Sampling of the survey, merging and linking of files, interpretation of missing values, methodology of the survey, unique ids, constructed variables etc. These are all given as part of the user guide.

(Refer Slide Time: 28:41)

Constructed Variables: Case and Geographic Identification

IHDS households are uniquely identified by the combination of stateid + distid + psuid + hhid + hhsplitid (where "+" signifies concatenation, not addition). In the individual file, persons are uniquely identified by the combination of those variables + personid. Several other identification variables are available to assist in sorting and merging files and for identifying geographic areas.

| Variable | Obs | Unique | Mean | Min | Max | Label |
|-----------|-------|--------|-------------|----------|-----------|--------------------------------|
| stateid | 41554 | 33 | 18.78 | 1 | 34 | State code |
| distid | 41554 | 61 | 14.69 | 0 | 68 | District code |
| psuid | 41554 | 39 | 5.76 | 1 | 39 | PSU: village/neighborhood code |
| hhid | 41554 | 52 | 9.22 | 1 | 52 | Household ID |
| hhsplitid | 41554 | 8 | 0.41 | 0 | 7 | Split household ID |
| caseid | 41554 | 41554 | . | . | . | HH id: 11 byte string |
| idhh | 41554 | 41554 | 181680422.0 | 10201010 | 340006150 | HH id 9-digit unique |
| idpsu | 41554 | 2474 | 189288.0 | 10201 | 340006 | PSU id 6-digit unique |
| stateid2 | 41554 | 22 | 483.18 | 101 | 733 | State codes, collapsed |
| distname | 41554 | 373 | 1892.80 | 102 | 3400 | District codes with names |
| dist01 | 41554 | 61 | 14.67 | 0 | 68 | H1sp: District ID Census 2001 |
| urban | 41554 | 2 | 0.36 | 0 | 1 | Census: 2001 village/town |
| metro6 | 4133 | 6 | 2.97 | 1 | 6 | Largest 6 metro areas 1-6 |
| sweight | 41554 | 1526 | 4623.48 | 220.0 | 308216.4 | Design weights |

The constructed variables like case and geographic identification; so all those details are given. I think IHDS households are uniquely identified; this will explore in detail, how it is uniquely identified and which variables it requires. It is a combination at this moment I need not explain in detail; state id, district id, psu id, household id, household split id. If you just combine all those together, we may define that particular id; it is a combined id can be uniquely identified and can be tracked in different file. Can be merged, it will be helpful for merging different files for our analysis.

Whereas, in the individual file persons are uniquely identified by the combination of those variables. I know that you might be having some doubts on what you mean by uniquely identified variable, or the particular variable called uniquely identified. How is it so? So we will explain it with the particular command in our respective lecture. You may take a hold on it, so later it will be clarified for sure.

(Refer Slide Time: 29:51)

IHDS: A PANEL DATASET

- ❑ IHDS-2 re-interviewed about **83%** of the IHDS-1 **households plus any split households** that resided in the same community.
- ❑ In order to link, Linking files are required (can be downloaded from IHDS website)

Data Download From IHDS

- Indian Paradox: Rising Education, Declining Women's Employment
- Ever-married Women Linking File - STATA Forma
- IHDS Individual Linking Variables-Text File
- IHDS Household Linking Variables-Text File
- IHDS Ever Married Women Linking Variables-Text File
- IHDS Individual Linking Variables-STATA format

28

So, IHDS is a panel dataset as I already said 83 percent re-interviews are there; Plus any split household that resided in the same community; community here we are referring to the village we already discussed. In order to link, linking files are required for making it a panel; I am not discussing in detail. This will be discussed later, how panel can be made.

(Refer Slide Time: 30:23)

- ❑ Recently, the panel version of IHDS data has been populated on DSDR website.

Web-link:
<https://www.icpsr.umich.edu/icpsrweb/DSDR/studies/37382>

- ❑ Linking file gives **Round-1 identification codes** for all Round-2 households that were re-interviewed at both the levels: Household and Individual.

29

So, panel version of IHDS data populated by IHDS very recently. It is on their website, you can have idea on it; so linking files give round-1 identification codes. For all round-2 households that were re-interviewed at both the levels; household as well as individual linking files.

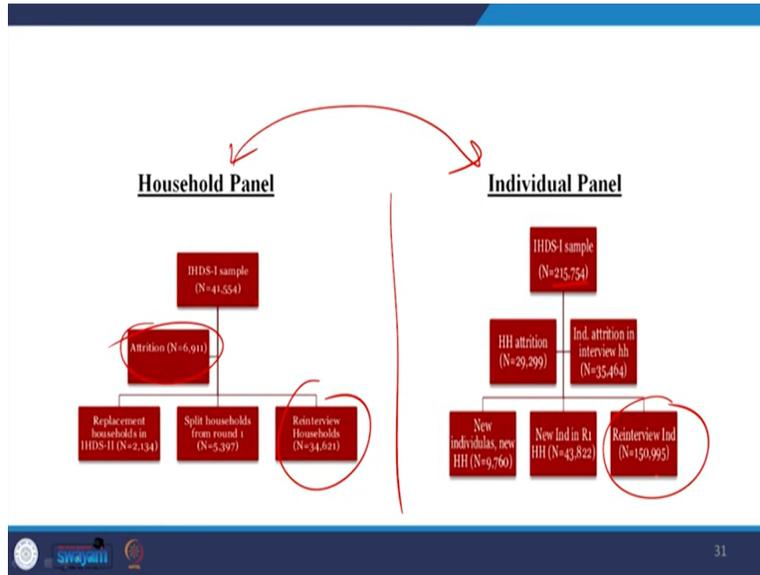
(Refer Slide Time: 30:43)

- Linking Variables: STATEID, DISTID, PSUID, HHID, HHSPLITID, PERSONID.
- Ever married women file will be merged with household file (no ever married file in IHDS-1).



Will discuss there are some doubts, I know that I am not clarifying and at this moment; because it is too heavy to be loaded. For you, will carry forward to the particular class how to make it panel; how to make it uniquely identified in the respective class.

(Refer Slide Time: 30:55)



But, there are two as I just said individual and household panel; this can be separately made. Household panel is with some attrition; there are some missing information. So, re-interviews are of 34,621; whereas in case of individual 1,50,995. Out of the total sample of 2,15,754; so 1,50,995 re-interviewed, there are also some attrition. So, we will discuss correctly in detail in the respective class. So, I think I should stop here. Friends we will discuss some of the further conceptual aspect to it in its respective class. With this let me stop here. Thank you.