

**Exploring Survey Data on Health Care**  
**Prof. Pratap C. Mohanty**  
**Department of Humanities and Social Sciences**  
**Indian Institute of Technology, Roorkee**

**Lecture - 23**  
**Merging Blocks of Information after Extraction**

Welcome friends, once again to this NPTEL MOOC module on Exploring Health Data Health Care Survey Data. We are on the week where we have been trying to explain the survey data analysis. We have been trained in handling these survey data. In the last year also, I handled this kind of explanation, this kind of survey data with the help of Stata.

Today also we are going to explain how survey data could be merged. On the very last lecture we have done the extraction of the survey data from the ASCII format. And we have guided you about 50-55 minutes for how you could have gone through from the scratch to the final shape of the data.

But still that was not complete, and this has been helped by many. This particular session we have been working for over 4-5 years and developed a systematic presentation for you. This has been supported by Mr. Milind, Mr. Kamal and Shruti. Shruti has already helped in the earlier last module.

So, you can have these contacts especially Kamal and Milind and you can be in touch with them through our NPTEL portal. They will be give you better guidance as well. So, here we are targeting to have merging of blocks of information after extraction. Now, why merging of blocks are required? I told you earlier that merging of blocks are essential because different blocks or levels give a different set up information.

So, you cannot be able to interpret or fine-tuned or a correct result just based on one block or one level of information. So, you have to have merge the essential files or essential labels. So, while you are going to merge the blocks some of the important steps, we are going to guide you through this session so that it becomes very easier for you to run those commands correctly and to process your data.

So, this is a clear-cut handholding session we are not keeping any theories here these are all practical and we will also operate through the software and where we could able to find out the right database before we start for our analysis.

So, here it goes. Starting in the introduction I just wanted to explain you that combining the dataset or the blocks is in fact called common data management task. And in many empirical projects the raw data to be utilized are in fact stored in separate files.

For example, separate files for household and individual records, separate files for each stages, separate webs of panel data, separate files for each category of information. Those are in fact called blocks of information as provided in the schedule of the survey.

(Refer Slide Time: 04:11)

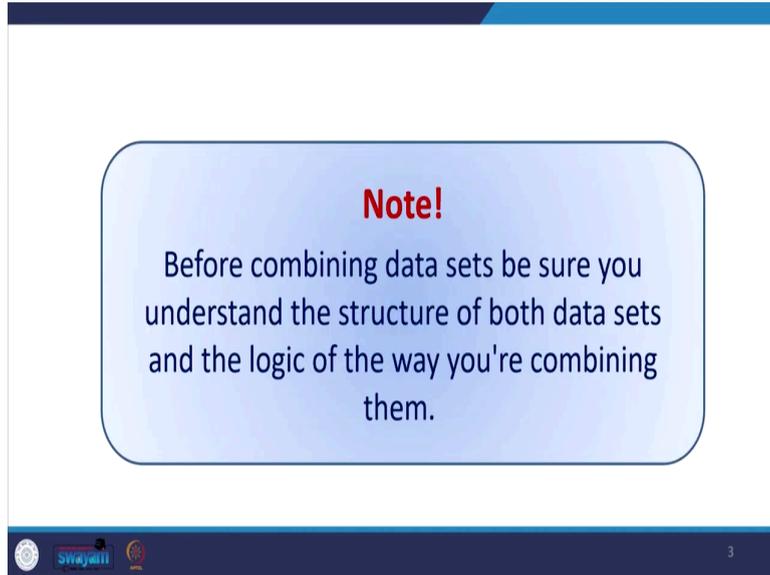
**Introduction**

- ❑ In many empirical research projects, the raw data to be utilised are stored in separate files. For example, separate files for household and individual records, separate files for each states, separate waves of panel data, separate files for each category of information (blocks in the questionnaires) etc.
- ❑ Combining two datasets is a common data management tasks. **Merging** and **appending** are such tools that combines two datasets.

2

So, combining is essential, merging and appending are such tools that combines two data sets.

(Refer Slide Time: 04:23)



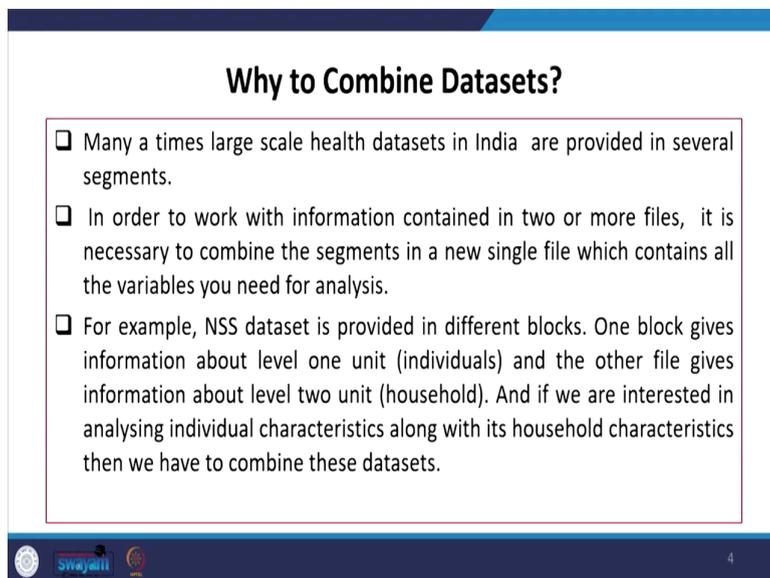
**Note!**

Before combining data sets be sure you understand the structure of both data sets and the logic of the way you're combining them.

swayam 3

From the beginning we are keeping a very strong note which will be very helpful for you to take enough care before combining. So, before combining data sets be sure you understand the structure of about data sets and the logic of the way you are trying to combine all of them.

(Refer Slide Time: 04:44)



**Why to Combine Datasets?**

- ❑ Many a times large scale health datasets in India are provided in several segments.
- ❑ In order to work with information contained in two or more files, it is necessary to combine the segments in a new single file which contains all the variables you need for analysis.
- ❑ For example, NSS dataset is provided in different blocks. One block gives information about level one unit (individuals) and the other file gives information about level two unit (household). And if we are interested in analysing individual characteristics along with its household characteristics then we have to combine these datasets.

swayam 4

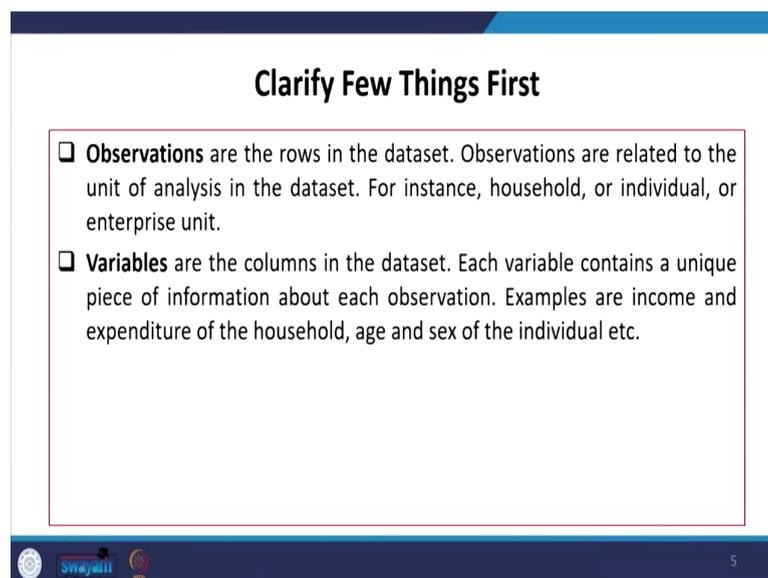
So, why to combine data set I have already mentioned. We all know that these days in the large-scale data sets is coming up with several structural components. And each of the component gives separate information. And they contain two or more files as well, it is

necessary to provide segments in a new single file which contains all variables you need to analyze.

For example, in the NSS data set as provided in different rounds it has different blocks. One block gives one setup information maybe for individual information another one is on household information, if we are interested in understanding the household characteristics as well as the individual characteristics.

So, that my analysis is going to be somewhere addressing the requirement. So, I have to combine these data sets. There are few clarifications we must do first then we go for steps. Firstly, we need to know which are called the observation and which are called the variables. Observations are the rows in the data sets like in a Excel data set if you remember or in the Stata spreadsheet.

(Refer Slide Time: 06:11)



**Clarify Few Things First**

- ❑ **Observations** are the rows in the dataset. Observations are related to the unit of analysis in the dataset. For instance, household, or individual, or enterprise unit.
- ❑ **Variables** are the columns in the dataset. Each variable contains a unique piece of information about each observation. Examples are income and expenditure of the household, age and sex of the individual etc.

swjani 5

Stata browse window you will get the data set, in which there are a columns of information and a rows of information. The rows in fact, are counterred as number of observations or the respondents in any survey. So, these observations are related to the unit of analysis in the data sets, for instance we have household or individual or enterprise units.

Variables are kept in the columns of that data set each variable contains a unique place of information about each observation. Example such as like income, expenditure of the household, sex and age of the individual etc.

(Refer Slide Time: 06:55)

	EDFID	CentreCode-d	Vill_Bla_fline	Round	Schedule	Sample	Sector	_sector	State_Region	State	_w
1	4383911002	718	43605	73	234	1	1	rural	091	09	
2	838391101	710	43639	73	234	1	2	urban	011	08	
3	614651401	717	61144	73	234	1	2	urban	212	21	
4	477931101	718	47793	73	234	1	1	rural	243	24	
5	474891403	718	47489	73	234	1	1	rural	241	24	
6	877741403	700	87774	73	234	1	2	urban	272	27	NA
7	6519310702	710	65193	73	234	1	2	urban	091	09	
8	484151404	717	48415	73	234	1	1	rural	351	35	
9	638541403	710	63854	73	234	1	2	urban	041	04	
10	651211002	710	65121	73	234	1	2	urban	093	09	
11	654821404	710	65482	73	234	1	2	urban	092	09	
12	940971401	710	94097	73	234	1	1	rural	321	32	
13	6193810201	717	61938	73	234	1	2	urban	193	19	
14	887471402	717	88747	73	234	1	1	rural	142	14	
15	9430711002	718	94307	73	234	1	1	rural	333	33	
16	9419310201	718	94193	73	234	1	1	rural	341	34	
17	6388911404	710	63889	73	234	1	2	urban	041	04	
18	882341001	717	88234	73	234	1	2	urban	194	19	
19	494471101	710	49447	73	234	1	1	rural	091	09	
20	673221401	718	67322	73	234	1	2	urban	242	24	
21	479421403	718	47942	73	234	1	1	rural	241	24	
22	4885910101	710	48859	73	234	1	1	rural	111	11	
23	6136910101	717	61369	73	234	1	1	rural	232	23	
24	437911002	710	43791	73	234	1	1	rural	012	01	
25	8449411002	710	84494	73	234	1	2	urban	094	09	

So, here we have kept for your clarity further on the data set window. We have marked as which are called observations on the row and on the column they are the variables.

(Refer Slide Time: 07:09)

### Two Ways of Combining Datasets

- ❑ **Merge** : adds the **different variables** to a dataset by joining two datasets together (more or less the same observations).
- ❑ **Append** : adds the **different observations** to a dataset by joining two datasets together (more or less same the same variables).
- ❑ In merge, we join two datasets :
  - Master file : the data file with which we will merge the other file.
  - Using file : the data file we will be merging with the master file.
- ❑ Stata always works with one dataset in the memory. So combining datasets here means combining the dataset in the memory (**master file**) with another dataset on disk (**using file**).

There are two ways of combining data sets. We are coming into the depth of the analysis now. We are now going into the deeper aspects of those combining data sets. So, there are broadly two ways of combining; one is called merging another one is called appending.

So, when we try to take the steps of merging or we try to do it with merging, we are actually adding the different variables into the number of observations. So, more information's we are actually adding more variables to a data set by joining two data sets together.

So, more or less the same observation, not necessarily the same observations are again carried forward rather maybe little more or even little less, but for sure we are adding more variables. Whereas in case of append we are actually adding observations, to the data set; variable maybe same or maybe little different.

Usually, we carry with the same variables and adding the observations. In merging we join two data set one is called master file and another is called using file, I am going to use it and then discuss with you very shortly about this in detail.

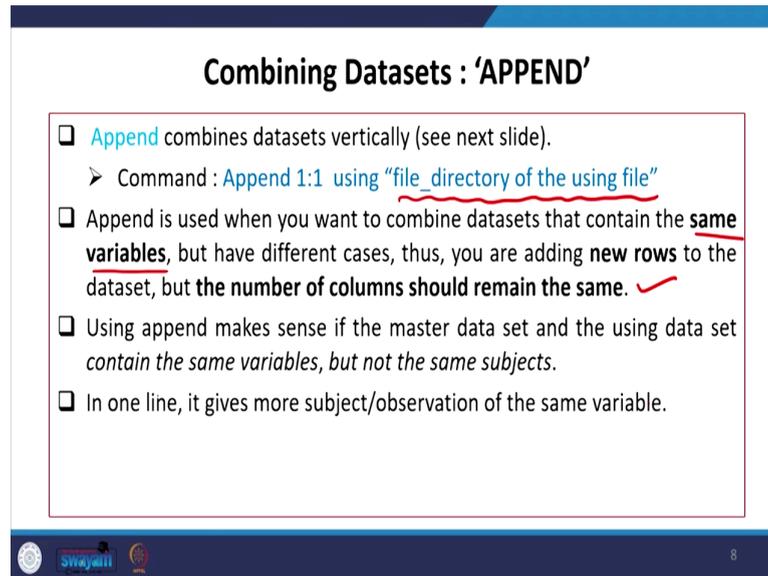
Master files are the files which we first open and that carries our beginning file. The data file with which we will merge the other file, which we are going to add to that file is called master file. So, using file which are to be added, that the file will be merging with the master file.

These you have to remember, Stata always works with one data set in the memory, but I think earlier in the or earlier versions of Stata, it gives two files together two window can be opened, but nowadays it is only with one memory.

So, combining data sets here means you are actually combining the data set in the memory. So, memory means you have already kept in the memory as the master file and you are in fact adding the using file to the memory. That is how Stata works for merging.

Regarding append data, it is all about combining data set vertically like observations are actually added, then that is why we say they are vertically added. The command we usually go for this is called append.

(Refer Slide Time: 10:06)



**Combining Datasets : 'APPEND'**

- ❑ **Append** combines datasets vertically (see next slide).
  - Command : **Append 1:1** using "file\_directory of the using file"
- ❑ Append is used when you want to combine datasets that contain the **same variables**, but have different cases, thus, you are adding **new rows** to the dataset, but **the number of columns should remain the same.** ✓
- ❑ Using append makes sense if the master data set and the using data set *contain the same variables, but not the same subjects.*
- ❑ In one line, it gives more subject/observation of the same variable.

swayam 8

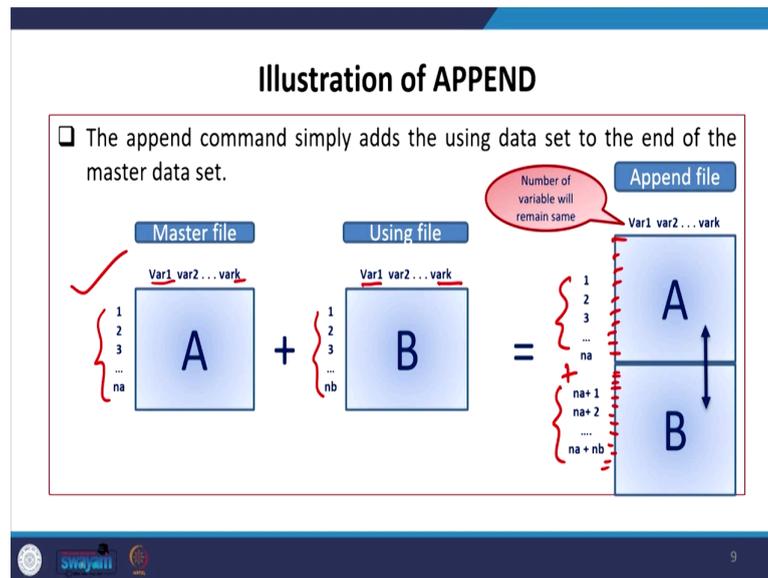
Append 1:1 using file directory basically what is this file director of the using file. Wherever you have the using file you give the file directory with invited comma, this is path name should be kept here whichever you wanted to a add. So, 1:1 basically all the variables are going to be added (to be carried forward) with the observations.

So, append is used when you want to combine data sets that contain the same variables. That is what I am saying when you have 1:1; that means, you have the same number of variables you basically wanted to add the number of observation to the corresponding same set of variables.

If you have different cases and you wanted to add then append command is the right approach. Thus, you are adding new rows to the data set, but the number of columns should remain the same. So, this is very clearly noted. Using append makes sense if the master data set and the using data set contain the same variable, but not the same subject.

So, subjects basically the observation should be different, but the variables should be same, so that append is actually considered. In one line it gives more subject or observations to the same variable. So, basically you are going to get the same variable, but more observations are added.

(Refer Slide Time: 11:45)



We will also show it here with the help of diagram. Append is explained with this figure like first box is for master file, second one is a using file. Once you have opened this we wanted to in fact add the B i.e., we wanted to add the information of B.

What is important to note the variable are same till V variable k, these are same. Now, your number of observation is still here, it is still na, now it is from another set of observations, but their responses with the variables name are same. So, in case of append we are actually adding all those extra variables.

We are just keeping like till na these are all entries observations we are entering the observation on this rows till this. So, these are called vertical addition, appending here refers to vertical addition of observations.

(Refer Slide Time: 12:53)

**Some preconditions of append:**

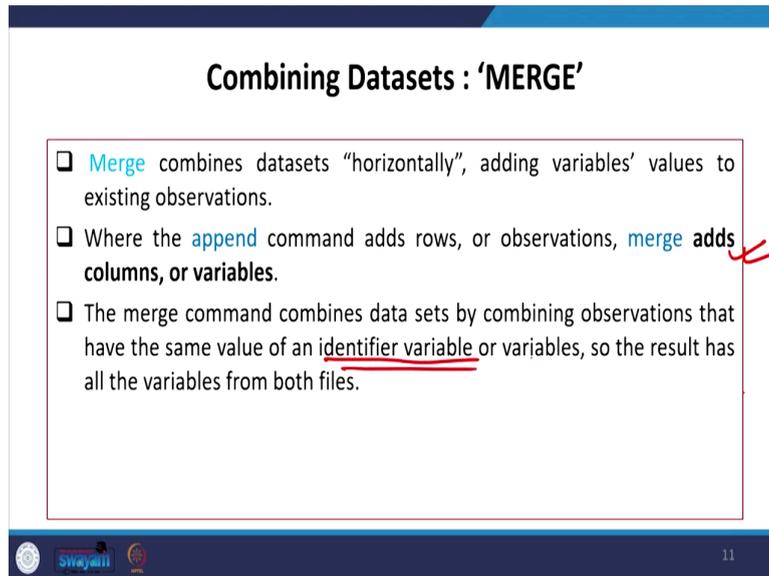
- The datasets should have most or all of the **same variables**. Make sure variables have the **same name** in both files before appending them, or append will treat them as different variables.
- Variables not common to all datasets can be appended. If a variable only exists in one of the two data sets, observations from the other data set will have **missing values** for that variable.
- Variables with the same name should have the same storage type (string, float, etc.).

So, some preconditions to append like the dataset should have most or all of the variables, make sure variables have the same name in both the files before appending them. The append will treat them as different variables or if you have different variable, then it will be treating them different of variables.

Variables not common to all data set can be appended, if a variable only exists in one of the two data sets observations from the other data set will have missing values for that particular variable. So that means, if two data set A and data set B, if the variable 1 and variable 2 if you are just going to add it. But like these does not have variables for v 2, but this one it has.

So, if you are appending it these will be coming out to be the missing values, so missing dots will be visible after appending. Variables with the same name should have the same storage type, so that is another most important aspect, should take a note of it that the variables should have the same storage as well as otherwise it will create some problem. So, string value or float value etc. if variables on string then the using file must have also string variable.

(Refer Slide Time: 14:27)



**Combining Datasets : 'MERGE'**

- ❑ Merge combines datasets “horizontally”, adding variables’ values to existing observations.
- ❑ Where the **append** command adds rows, or observations, **merge adds columns, or variables.**
- ❑ The merge command combines data sets by combining observations that have the same value of an identifier variable or variables, so the result has all the variables from both files.

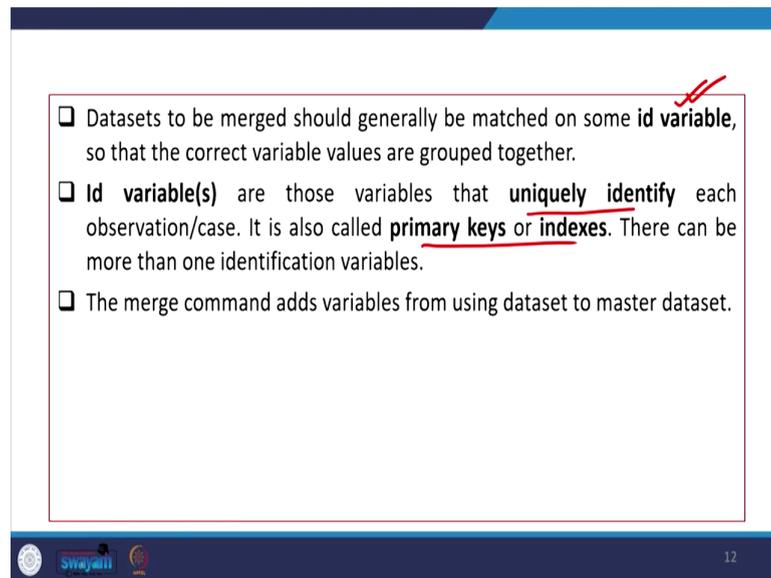
11

Combining data set that is merging is most important in NSS data particularly and other data as well. All unit level data merging is very essential. Merge combines data sets horizontally and adding variable values to their existing observation.

Where the append command adds rows or observations and merge adds columns or variables, that is in fact another important distinction. The merge command combines data sets by combining observation that have the same value of an identifier variable or variable, so that the result has all the variables from both the files.

So, identifier variables are important. How identifier variables are actually considered we are going to explain very shortly.

(Refer Slide Time: 15:22)



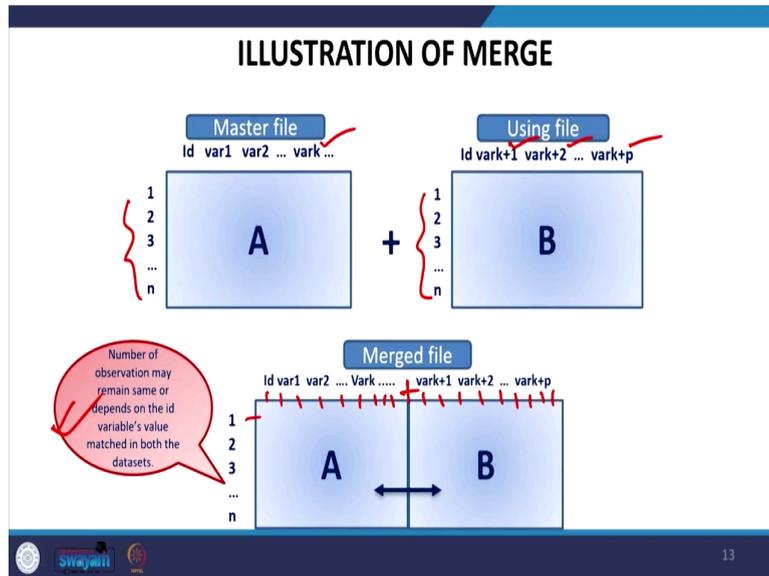
- Datasets to be merged should generally be matched on some **id variable**, so that the correct variable values are grouped together.
- Id variable(s)** are those variables that uniquely identify each observation/case. It is also called primary keys or indexes. There can be more than one identification variables.
- The merge command adds variables from using dataset to master dataset.

So, data sets to be merged should generally be matched on some id variable, this is what we are trying to say about the unique id variable or the primary keys. So, that the correct variable values are grouped together. Id variables are those variables that that uniquely identify, each observation and it is also called primary keys or indexes; there can be more than one identification variables.

What does this mean primary key explains that the once a key is assigned in each of the blocks, if you have to just merge that particular key you are actually pulling other information corresponding to that particular primary key.

So, the primary key is actually going to combine you both the data sets with right information. The merge command adds variables using data sets to master data sets like this.

(Refer Slide Time: 16:22)



So, master data set is A here using data set is B. Here number of observation may be same, but the variables are actually different this is variable till vk this is vk plus 1, 2 until p. So, all will be added, so this is variable entries, here we are actually adding new variables corresponding to each of the observations.

Number of observation may remain same or depends on the id valuable values. So, if the id variable values are perfectly corresponding or perfectly identify in both the case by 1:1, then number of observation is expected to be the same.

(Refer Slide Time: 17:13)

### Key to Merging

- ❑ Like the append command, the `merge` works on a “master” dataset—the current contents of memory—and a single “using” dataset.
- ❑ The data sets to be merged must all contain at least one “linking” variable that has the **same name** in each data set. And, it must **uniquely identify** each subject id.
- ❑ There can be more than one unique id variable.
- ❑ The other variables should be different in the two data sets.
- ❑ You must sort each dataset by the id variable prior to their merge.

So, key to merging there are certain important aspects like the append command the merge works on a master data set, the current contents of the memory and a single using data set. Like master and a single using data set to be merged. The data sets to be merged must all contain at least one linking variable, this is another important aspect which is essential. Another is that the same name in both the data set, then only both the data sets are expected to be merged.

And it should actually uniquely identify each of the subject id. There can be more than one unique id variable, the other variable should be different in two data sets. You must sort each data set by id variable prior to their merge. So, another step we are going to guide you right now you must also sort this database. Then only it is going to merge the different variables.

So, if the corresponding unique id's are not in a systematic order, the data cannot able to actually recognize in another using file. So, if it is sort; that means, it the second file is also sorted very correctly. That means, while you merge you can be able to get the right corresponding information.

(Refer Slide Time: 18:43)

The slide is titled "Process of Merge" and contains the following content:

- ❑ The syntax for merge is simple. Start by opening the master dataset.  
`merge (1:1/ 1:m/ m:1/ m:m) id variable using "merging dataset"`
- ❑ In theory, there are 4 kinds of merges:
  - One to one (1:1),
  - One to many (1:m),
  - Many to one (m:1), and
  - Many to many (m:m)

The slide also features a red bracket grouping the last three merge types and a Swayam logo at the bottom left.

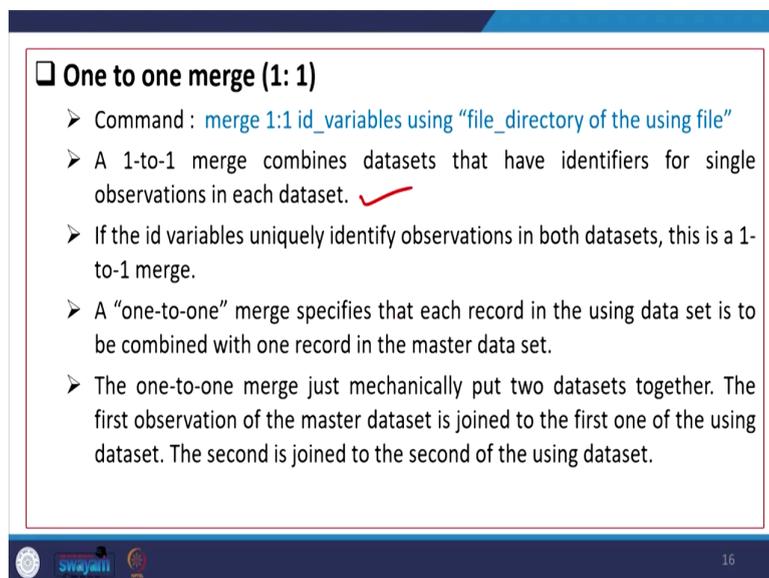
Regarding the steps for merging, there are four important steps we usually follow they are called 1:1, 1:m, m:1 and m:m. So, these are like 1:1 stands for 1 to 1, merging 1 is to m for 1 to many, then many to 1 then many-to-many; so these are four guidance.

Now, we need to start by opening the master data set first, I will show you all those practical aspects on our Stata software. So, we will start by opening the master data set, master data set will open any data set we first open. Then if you are thinking of adding another file to the master data set, that data set is in fact called using data set whichever you are going to use it. But the first one which we have opened is in fact called the master one.

So, the command we use is called merge if it is 1:1. I will clarify all those detail in between 1:1 or 1 is to any one options to be picked up. Let it be if it is corresponding perfectly with 1:1, 1 are perfectly matching by their unique id; that means, they are actually 1:1. If all are not matching in the using file only first one is each of them uniquely identified, but others are not then in that case if they have multiple information then it will be 1:m, m stands for multiple information.

The reverse you can also do it multiple information to 1. Usually, we follow the first two methods others are not suggested much.

(Refer Slide Time: 20:35)



**One to one merge (1: 1)**

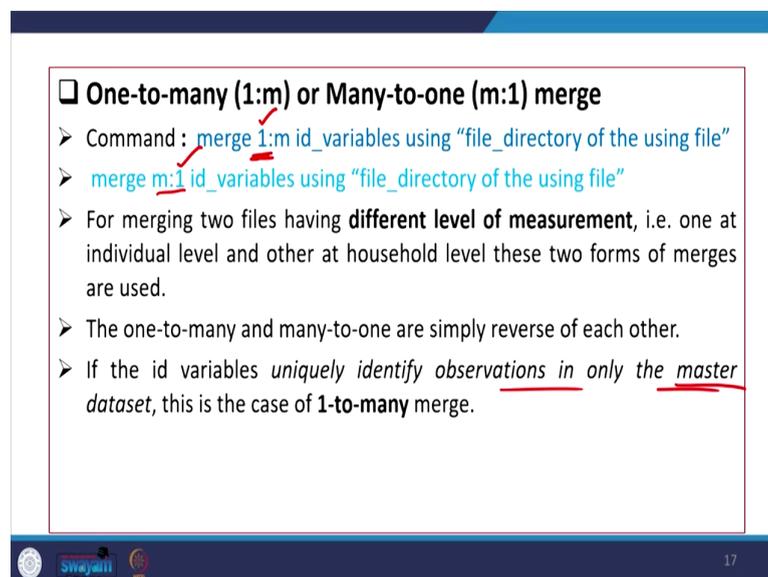
- Command : `merge 1:1 id_variables using "file_directory of the using file"`
- A 1-to-1 merge combines datasets that have identifiers for single observations in each dataset. ✓
- If the id variables uniquely identify observations in both datasets, this is a 1-to-1 merge.
- A "one-to-one" merge specifies that each record in the using data set is to be combined with one record in the master data set.
- The one-to-one merge just mechanically put two datasets together. The first observation of the master dataset is joined to the first one of the using dataset. The second is joined to the second of the using dataset.

So, like what do you mean by 1 is to 1 merging: this is basically the command where both the files i.e., master file as well as the using file have perfect identifiers for each single observation in the data set. Each single data set have the same identifier and each one is the individual kind. Like perfectly one each identify with another perfectly with their observations, then that case we apply 1 is to 1 merging.

If the id variables uniquely identify observation in both the data set that is basically called 1 is to 1 merge. A 1-to-1 merge specifies that each record in the using data set is to be combined with one record in the master data set. The 1-to-1 merge just mechanically put two data sets together, the first observation of the master data set is joined to the first one of the using data set, the second is joined to the second one from the using data set.

So, that is why what it follows, the rest you will be guided through our practical session here after couple of minutes. Second one is called one is to many I am going to use all those techniques one is to many or on the reverse called many is to one merge. So, 1 is to many we need to cite give the id variables first then using file.

(Refer Slide Time: 22:09)



**One-to-many (1:m) or Many-to-one (m:1) merge**

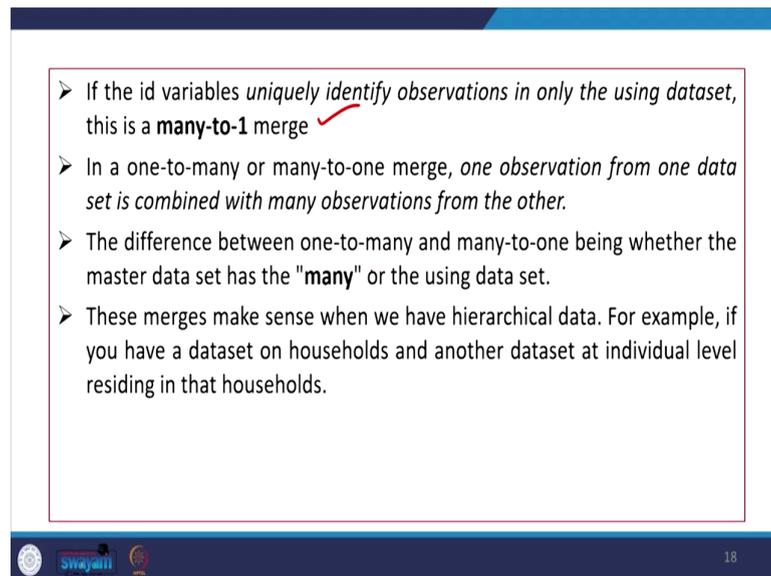
- Command : `merge 1:m id_variables using "file_directory of the using file"`
- `merge m:1 id_variables using "file_directory of the using file"`
- For merging two files having **different level of measurement**, i.e. one at individual level and other at household level these two forms of merges are used.
- The one-to-many and many-to-one are simply reverse of each other.
- If the id variables *uniquely identify observations in only the master dataset*, this is the case of **1-to-many** merge.

Then we will go by the using file directory, basically we need to specify the exact, directory the path name of the particular file. Similarly, if we go by the reverse m is to 1, then id variables of the using file should be actually specified.

If it is like 1 is to m then the first file has to be uniquely identified, that is the master file the master files must be uniquely identified. If it is the second file that is the using file using file must be uniquely identified for sure then it works very correctly. For merging two files having different levels of measurement like one at individual level and other at the household level.

If that is the case in these two files what is going to happen the 1 is to many or and many is to 1 are simply reverse of each other. If the id variables uniquely identify observations in only the master of data set that is where you are supposed to apply the 1 is to many merge in command. If the master file is only uniquely identified then 1 is to since your master file is actually 1 and others may not be uniquely identified.

(Refer Slide Time: 23:47)



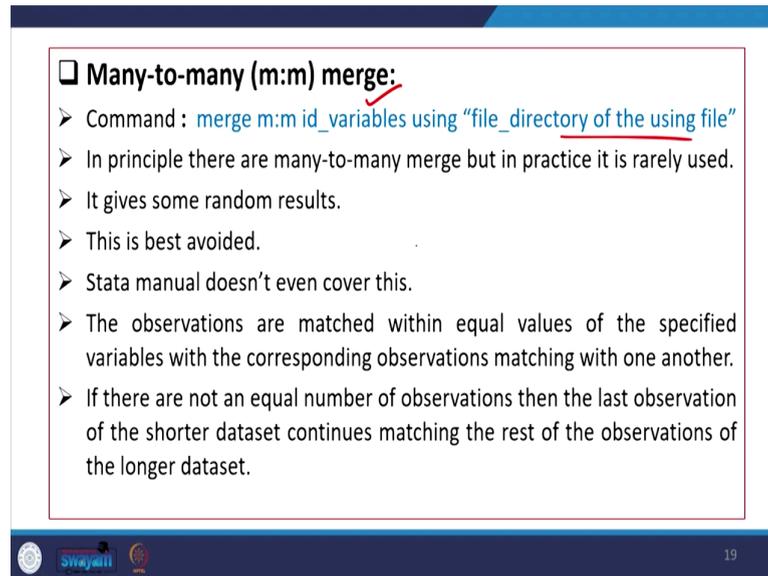
- If the id variables *uniquely identify observations in only the using dataset*, this is a **many-to-1** merge ✓
- In a one-to-many or many-to-one merge, *one observation from one data set is combined with many observations from the other*.
- The difference between one-to-many and many-to-one being whether the master data set has the "**many**" or the using data set.
- These merges make sense when we have hierarchical data. For example, if you have a dataset on households and another dataset at individual level residing in that households.

If the id variables uniquely identify observations in only the using data set then many is to 1 command is perfectly suitable. So, these type of questions might be asked in the exam that which one you apply the most, we will give all the options MCQs options. So, in 1 to many or many to 1 merge, one observation from one data set is combined with many observation from the other.

1 to many means 1 is combined with many information; that means, the file you are expected to be much larger than the first one and many observations are going to be added. The difference between 1 to many and many to 1 being whether the master data set has the many or the using data set, where we have that many information that is important and accordingly we can take the steps.

These merges make sense when we have hierarchical data for example if you have a data set on households and another data set at individual level residing in that household.

(Refer Slide Time: 24:58)



**❑ Many-to-many (m:m) merge:**

- Command : `merge m:m id_variables using "file_directory_of_the_using_file"`
- In principle there are many-to-many merge but in practice it is rarely used.
- It gives some random results.
- This is best avoided.
- Stata manual doesn't even cover this.
- The observations are matched within equal values of the specified variables with the corresponding observations matching with one another.
- If there are not an equal number of observations then the last observation of the shorter dataset continues matching the rest of the observations of the longer dataset.

swayamii 19

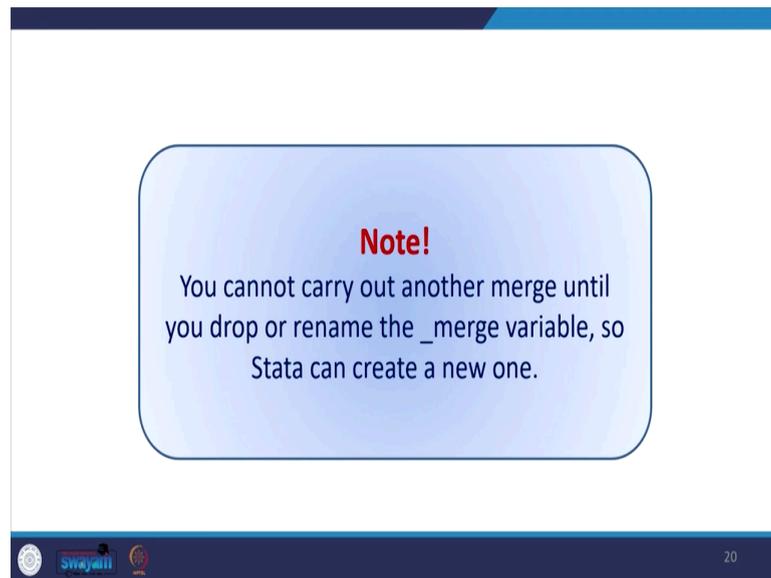
So, many-to-many merge command here, if many-to-many as I already guided just before you need not require the id variable to be uniquely identified. Since both the files are many-to-many, so since it is not having any one or each observations are not uniquely identified.

So, many-to-many with the id variables should be given, but id variables may not be uniquely identified. So, using the data file, so accordingly we can do that, I will apply just now through the data set. The sample data set we will also upload in your NPTEL portal.

In principle there are many-to-many merge, but in practice this is rarely used it gives some random results and this is suggested to avoid. Stata manual does not even cover this particular aspect the observations are matched with equal values of the specified variables with the corresponding observation matching with one another.

If there are not an equal number of observation then the last observation of the shorter data set continues, matching the rest of the observation of the longer data set. So, anyway that gets merged.

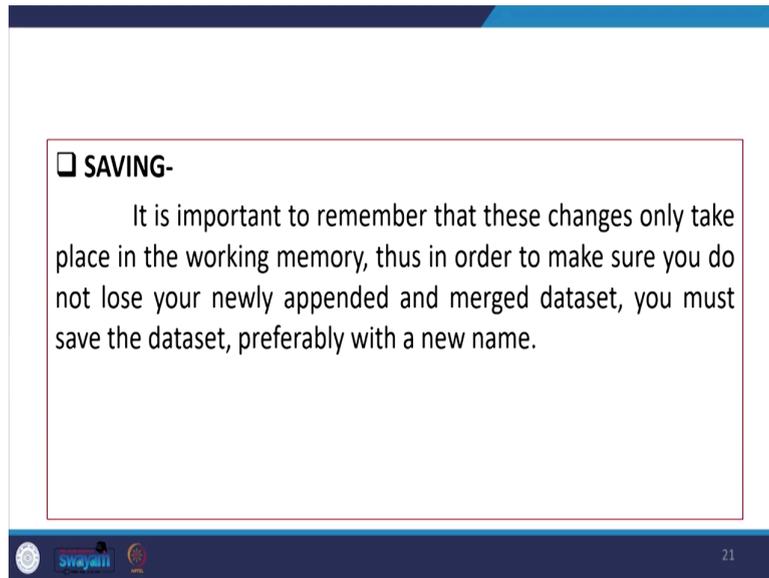
(Refer Slide Time: 26:21)



So, one note, we wanted to just pass it to you that you cannot carry out another merge until you drop or rename the merge variable, so Stata can create a new one. So, you have to actually rename once you have merge it you will come up with this kind of information in the variable window.

Underscore merge will come out and that has to be either renamed or this has to be drop in order to merge further. I will show those all those things you need to save also once you have merged. You need to save it is important to remember that the changes only take place in the working memory.

(Refer Slide Time: 27:16)



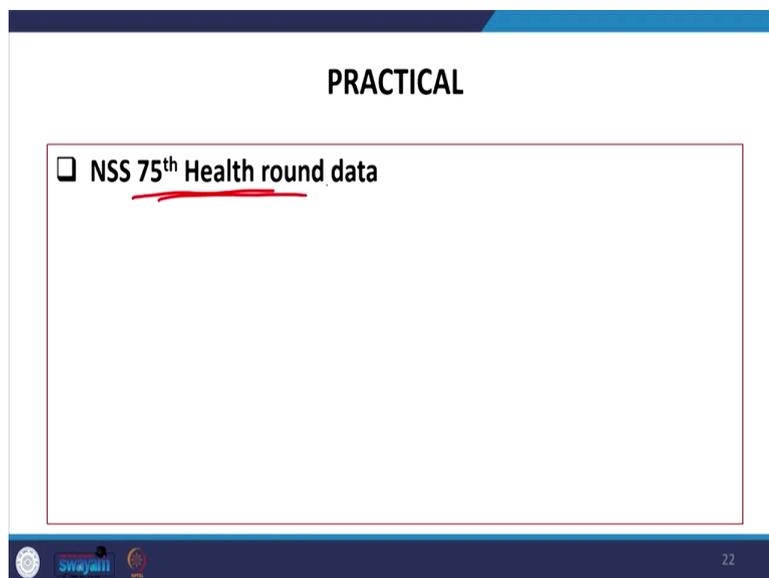
**SAVING-**

It is important to remember that these changes only take place in the working memory, thus in order to make sure you do not lose your newly appended and merged dataset, you must save the dataset, preferably with a new name.

21

Thus, in order to make sure you do not lose your newly appended and merged data set you must have to save the data set preferably with an appropriate new name.

(Refer Slide Time: 27:25)



**PRACTICAL**

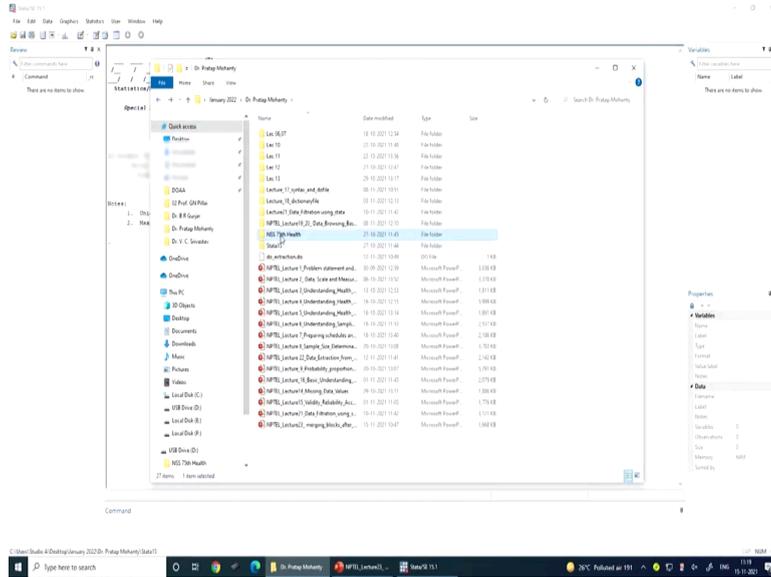
**NSS 75<sup>th</sup> Health round data**

22

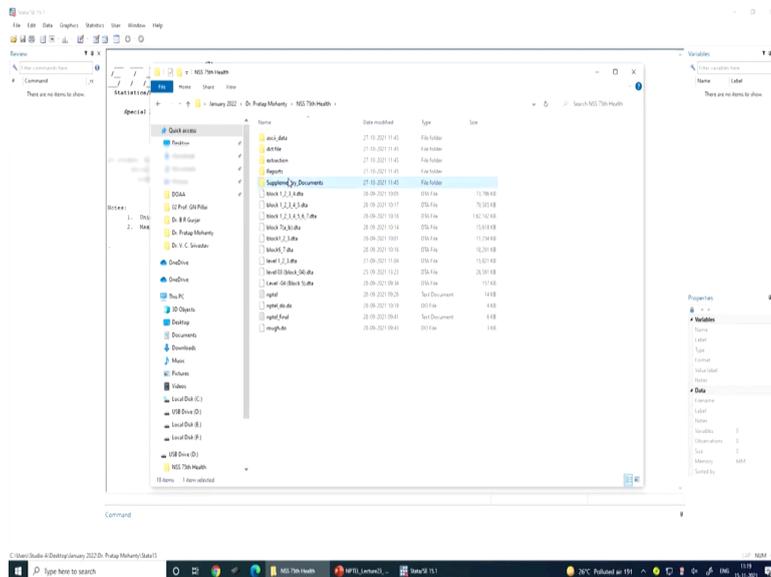
So, now we are going to show you the practical handouts or exercise using the NSS 75th round data.



(Refer Slide Time: 27:50)

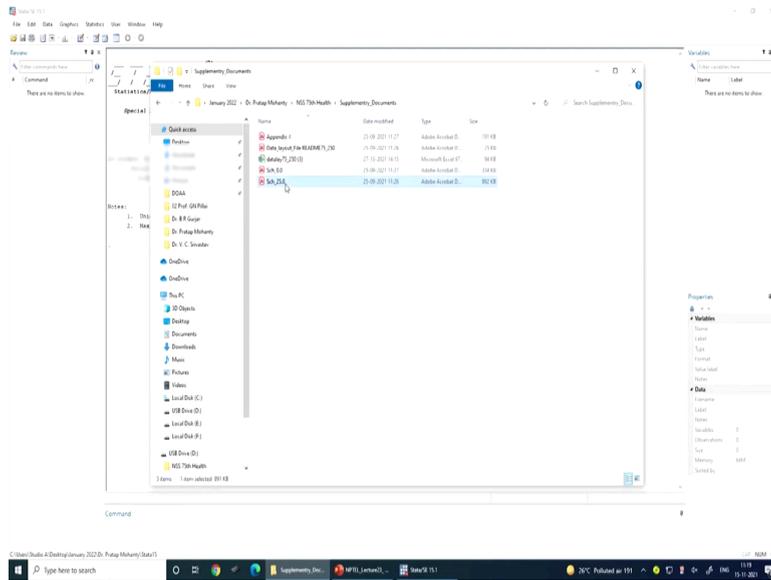


(Refer Slide Time: 27:53)



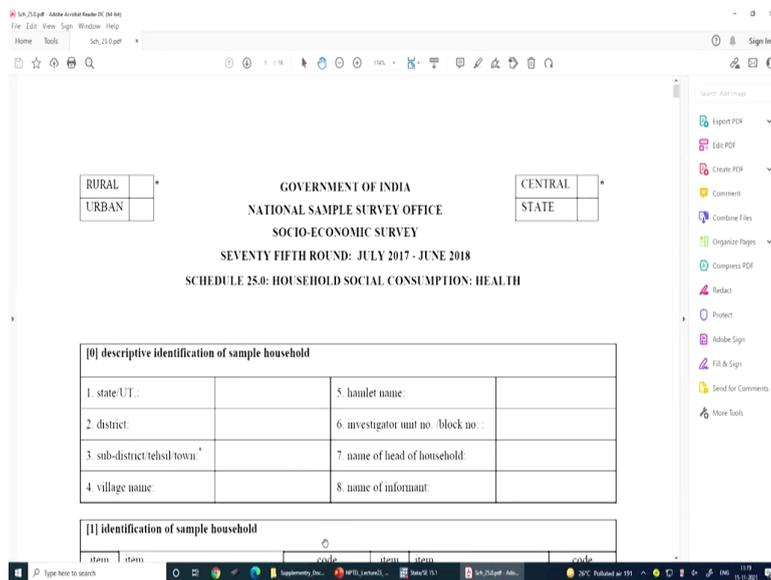
I will show you all those things together, 75th round we require layout file first.

(Refer Slide Time: 27:57)



So, we open schedule.

(Refer Slide Time: 28:00)



The schedule that is important, then layout file, read me file also we have opened.

(Refer Slide Time: 28:08)

**Government of India**  
**Data Quality Assurance Division**  
**National Statistics Office**  
**164, Gopal Lal Thakur Road, Kolkata-108.**  
**Phone No. 2577-1128**

-----  
**NSS 75<sup>th</sup> Round**  
**Final Multiplier-posted unit-level data**  
**for Schedule- 25.0 of NSS 75<sup>th</sup> round**

A) Data for Sch. 25.0 (Social Consumption: Health).  
 There are 13 data files belonging to 13 different levels as per layout (datalay75\_250.XLS).

File names	No. of Records
R75250L01.TXT	113823
R75250L02.TXT	113823

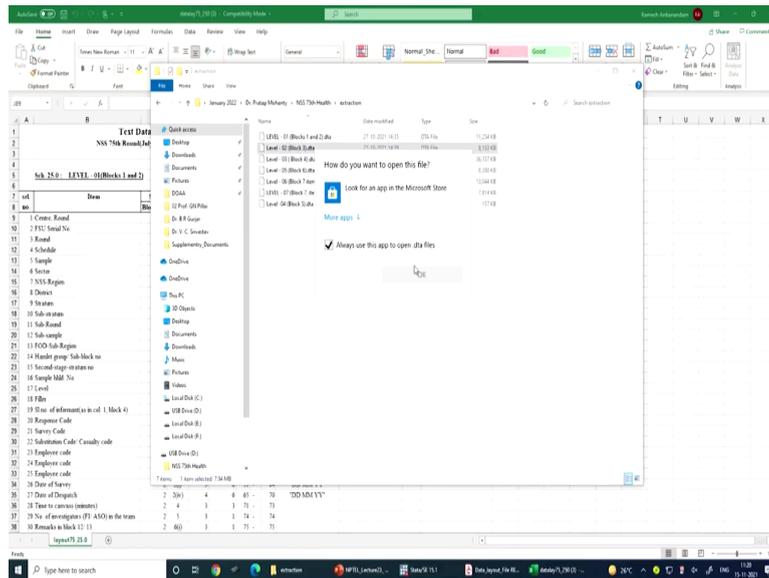
Here is the read me file then we will open layout file also and a micro data set.

(Refer Slide Time: 28:27)

Item	Schedule reference	Length	Remarks
1 Create Round		1 1 - 1	Generated
2 NSS Serial No	1 1	5 4 - 8	
3 Round	1 2	2 8 - 10	"00" Generated
4 Schedule	1 3	3 11 - 13	"00" Generated
5 Sample	1 14	14 Generated from Sub-ID	
6 Area	1 15	15 Generated from Sub-ID	
7 NSS-Region	1 16	16 Generated from Sub-ID	
8 District	2 19	19 Generated from Sub-ID	
9 Region	2 21	21 Generated from Sub-ID	
10 Sub-region	2 23	23 Generated from Sub-ID	
11 Sub-Round	1 25	25 Generated from Sub-ID	
12 Sub-sample	1 26	26 Generated from Sub-ID	
13 FOC Sub-Region	4 27	27 Generated from Sub-ID	
14 Handler group Sub-block no	1 4	3 31 - 31	
15 Second-stage-stratum no	1 5	1 32 - 32	
16 Sample-MSL No	1 6	2 33 - 34	"Common-00"
17 Level	2 35	2 35 - 36	"00" Generated
18 File	1 37	1 37 - 41	"0000" generated
19 Str no of information in col. 1, block 4	1 7	2 42 - 43	
20 Response Code	1 8	1 44 - 44	
21 Survey Code	1 9	1 45 - 45	
22 Substratum Code/ Casualty code	1 10	1 46 - 46	
23 Employee code	2 10(0)	1 47 - 50	
24 Employee code	2 10(0)	4 51 - 54	
25 Employee code	2 10(0)	3 4 55 - 58	
26 Date of Survey	2 7(0)	1 8 59 - 64	"00/00/00"
27 Date of Dispatch	2 3(0)	4 65 - 70	"00/00/00"
28 Time to census (minutes)	2 4	3 71 - 71	
29 No. of non-respondents (PT ASD) in the train	2 5	1 72 - 74	
30 Remarks in block 12 13	2 4(0)	1 75 - 75	

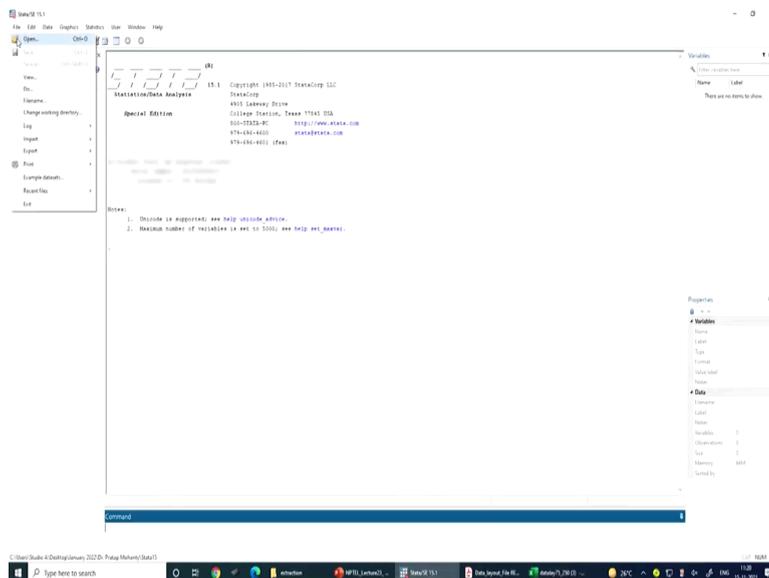
Now this is being opened. So, now, in the last lecture, I told you about extraction I have already explained extraction. So, now, we should have the extracted data files from the ASCII format. So now, I will explain all those things simultaneously.

(Refer Slide Time: 28:53)



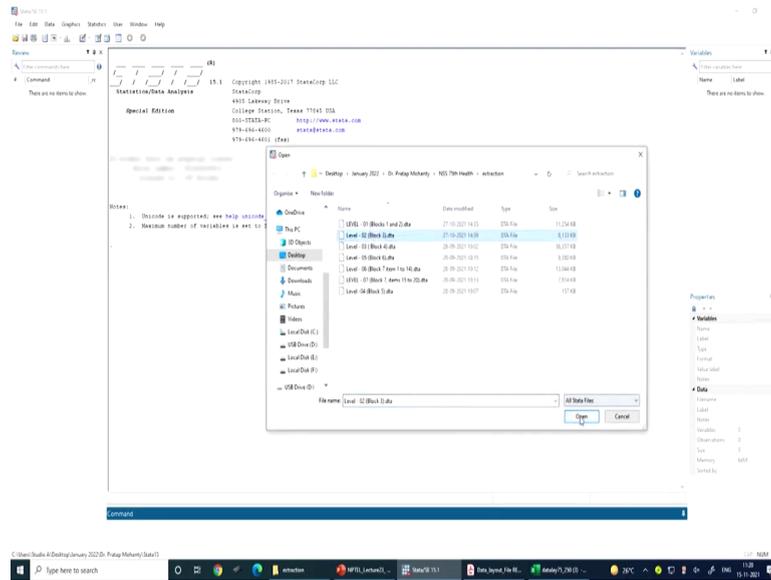
Let us open one extracted data, maybe level number 2.

(Refer Slide Time: 29:09)

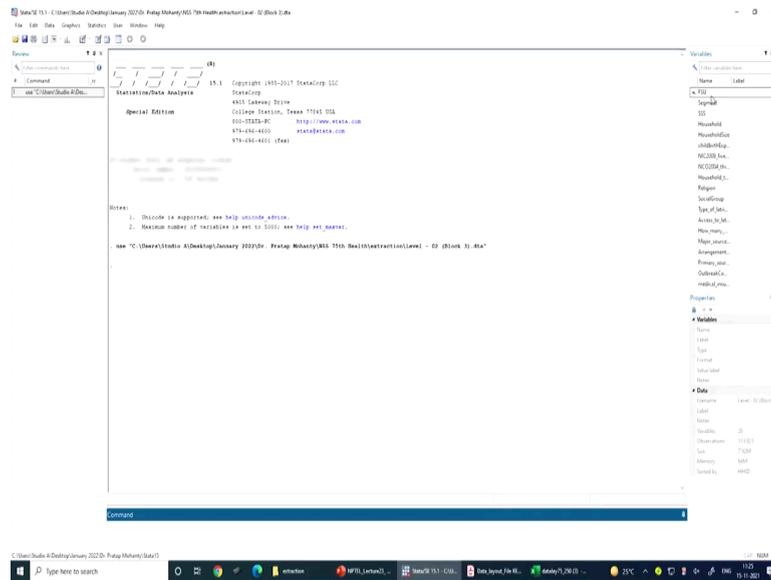


So, we are going to open the extracted file, that we did it in our last lecture.

(Refer Slide Time: 29:11)



(Refer Slide Time: 29:16)



So, on your window we have the level number 2 information. So, level number 2 as per the schedule we just go here.

(Refer Slide Time: 29:29)

File names	No. of Records
R75250L01.TXT	113823
R75250L02.TXT	113823
R75250L03.TXT	555352
R75250L04.TXT	2537
R75250L05.TXT	93925
R75250L06.TXT	93925
R75250L07.TXT	93925
R75250L08.TXT	43240
R75250L09.TXT	43240
R75250L10.TXT	43240
R75250L11.TXT	42762
R75250L12.TXT	70258
R75250L13.TXT	32257
Total	1342307

Record length for data is 142.

Here is the level number 2 it has the information about 1,13,823. This is what you can check it on the screen. Then second on level 2 you also have to check your schedule that which kind of variables are available on level 2, they are household type, religion social groups, type of latrine, access to latrine etc. These are called household information. These information are made for the household level not the individual information.

(Refer Slide Time: 30:17)

item no.	item	code	item no.	item	code	
1.	stl. no. of sample FSU		6	sample household number		
2.	round number	7	5	7.	serial number of informant (as in column 1 of block 4)	
3.	schedule number	2	5	0	8.	response code
4.	sample hg/sb number		9.	survey code		
5.	second-stage stratum number		10.	reason for substitution of original household (code)		

**CODES FOR BLOCK 1**

item 8 **response code:** informant: co-operative and capable -1, co-operative but not capable -2, busy -3, reluctant -4, others -9.

item 9 **survey code:** original -1, substitute -2, casualty -3.

item 10 **reason for substitution of original household:** informant busy -1, members away from home -2, informant non-cooperative -3, others -9.

\* tick mark (✓) may be put in the appropriate place

So, where these are located you can see in the schedule or the questionnaire file.

(Refer Slide Time: 30:20)

[1] particulars of field operations		Field Investigator (FI) / Junior Statistical Officer (JSO)			Field Officer (FO) / Senior Statistical Officer (SSO)		
sl. no.	item	(3)			(4)		
1 (a)	(i) name (block letters)						
	(ii) code						
	(iii) signature						
1 (b)	(i) name (block letters)						
	(ii) code						
	(iii) signature						
2	date(s) of:	DD	MM	YY	DD	MM	YY
	(i) survey inspection						
	(ii) receipt						
	(iii) scrutiny						
	(iv) despatch						
3	number of additional sheet(s) attached						
4	total time taken to canvass the schedule by the team of investigators (FI/JSO) (in minutes) [no decimal point]						
5	number of investigators (FI/JSO) in the team who						

(Refer Slide Time: 30:21)

4	total time taken to canvass the schedule by the team of investigators (FI/JSO) (in minutes) [no decimal point]						
5	number of investigators (FI/JSO) in the team who canvassed the schedule						
6	whether any remark has been entered by FI/JSO/supervisory officer (yes-1, no-2)	(i) in block 12/13					
		(ii) elsewhere in the schedule					

[12] remarks by investigator (FI/JSO)

[13] comments by supervisory officer(s)

Here is our questionnaire or schedule.

(Refer Slide Time: 30:23)

[3] household characteristics			
1	household size		8. type of latrine usually used (code)
2	whether the household paid major share for childbirth expenses for any non-household female member(s) during last 365 days? (yes-1, no-2)	If code in item 8 is 01-09	9. access to latrine: exclusive use-1, common use of households in the building-2, public community latrine-3, others-9
3	principal industry (NIC-2008)		10. how many members use the latrine?
4	description:		11. major source of drinking water (code)
	code (5-digit)		12. arrangement of garbage disposal (code)
4	description:		13. primary source of energy for cooking during the last 30 days (code)
	code (3-digit)		14. was there a sudden outbreak of communicable disease (see list* below) in the community afflicting at least one household member during last 365 days? (yes-1, no-2)
5	household type (code)		15. amount of medical insurance premium paid for household members during last 365 days (Rs.)
6	religion (code)		16. household's usual monthly consumer expenditure (Rs.)
7	social group (code)		

**CODES FOR BLOCK 3**

item 5: household type: for rural areas self-employed in agriculture -1, self-employed in non-agriculture - 2, regular wage/salary earning in agriculture- 3, regular wage/salary earning in non-agriculture- 4, casual labour in agriculture - 5, casual labour in non-agriculture -6, others-9

for urban areas self-employed -1, regular wage/salary earning - 2, casual labour -3, others -9

item 6: religion Hinduism -1, Islam -2, Christianity -3, Sikhism-4, Jainism -5, Buddhism -6, Zoroastrianism -7, others -9

item 7: social group Scheduled Tribes (ST)-1, Scheduled Castes (SC) -2, Other Backward Classes (OBC) -3, Others -9

item 8: type of latrine usually used: flush pour flush latrine to piped sewer system-01, septic tank-02, pit latrine-03, elsewhere (open area, street yard plot drainage ditch, not known etc.)-04  
Pit latrine ventilated improved pit, biogas latrine -05, pit latrine with slab-06, pit latrine without slab open pit-07, twin pit, composting latrine-08, others-09, no latrine facility use open space or field-10

item 11: major source of drinking water bottled water - 01, piped water in dwelling premises yard - 02, piped water outside -03, tube-well borewell (inside or outside premises)-04, protected well (inside or outside premises)-05, tanker truck/drum/supplied through container)-06, protected spring/pond etc for drinking purpose -07, Community RO Plant-08, others - 09, all unprotected source (river/canal, spring, pond) well etc 1 - 10

item 12: arrangement of garbage disposal: by Panchayati Municipality Corporation-1, by resident group of residents-2,

Yes, this is the one household characteristics. This is provided along with the NSS data which I already have guided in the last lecture and also in the very first week of our module where we have guided about NSS data. Here we said about household size household type, religion, social groups, then access to latrine etc. Their principal occupation, then principal industrial involvement etc., all those information are available.

(Refer Slide Time: 31:06)

**CODES FOR BLOCK 3**

item 5: household type: for rural areas self-employed in agriculture -1, self-employed in non-agriculture - 2, regular wage/salary earning in agriculture- 3, regular wage/salary earning in non-agriculture- 4, casual labour in agriculture - 5, casual labour in non-agriculture -6, others-9

for urban areas self-employed -1, regular wage/salary earning - 2, casual labour -3, others -9

item 6: religion Hinduism -1, Islam -2, Christianity -3, Sikhism-4, Jainism -5, Buddhism -6, Zoroastrianism -7, others -9

item 7: social group Scheduled Tribes (ST)-1, Scheduled Castes (SC) -2, Other Backward Classes (OBC) -3, Others -9

item 8: type of latrine usually used: flush pour flush latrine to piped sewer system-01, septic tank-02, pit latrine-03, elsewhere (open area, street yard plot drainage ditch, not known etc.)-04  
Pit latrine ventilated improved pit, biogas latrine -05, pit latrine with slab-06, pit latrine without slab open pit-07, twin pit, composting latrine-08, others-09, no latrine facility use open space or field-10

item 11: major source of drinking water bottled water - 01, piped water in dwelling premises yard - 02, piped water outside -03, tube-well borewell (inside or outside premises)-04, protected well (inside or outside premises)-05, tanker truck/drum/supplied through container)-06, protected spring/pond etc for drinking purpose -07, Community RO Plant-08, others - 09, all unprotected source (river/canal, spring, pond) well etc 1 - 10

item 12: arrangement of garbage disposal: by Panchayati Municipality Corporation-1, by resident group of residents-2,

(Refer Slide Time: 31:08)

Sl. no.	Name of member	Relation to head (code)	Gender (code)	Age (yrs)	Marital status (code)	Gen. edn. status (code)	Prin. activity (code)	during last 365 days									
								whether hospitalized (yes-1, no-2)	if in col. 9, no. of times hospitalized	whether pregnant* (yes-1, no-2)	whether paid major share for child-birth expense* (code)	whether suffered from any communicable disease (code)	whether suffering from any chronic ailment (yes-1, no-2)	whether suffered from any other ailment* (besides chronic ailments) on the day before the date of survey (yes-1, no-2)	whether covered by any scheme for health exp. support (code)	reporting of column 14-16 (col-1, col-2)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)

A: usual members

Now, on the next one is on the individual information like demographic particulars of household members. So, name of the member like NSS collects information about some of the members of the household with different serial number. What is relation to head, their gender, age, marital status, general education, principal activities of that particular person etc. is given. So, these are also figured in the layout file.

(Refer Slide Time: 31:42)

Item	Schedule reference	Length	Remarks
1. Census ID	14	1 - 14	Item-Appendix
2. Level	2	15 - 16	"02" General
3. File	3	17 - 41	"00007" General
4. Household size	1	2 - 42 - 43	
5. Whether HHID paid major share for child/birth expenses for any non-HHID	3	2 - 44 - 44	
6. NSO 2008 for age code	3	3 - 45 - 49	
7. NSO 2008 for age code	3	4 - 50 - 52	
8. Household type	3	1 - 53 - 53	
9. Religion	3	4 - 54 - 54	
10. Social group	3	7 - 55 - 55	
11. Type of latrine usually used	3	2 - 56 - 57	
12. Access to latrine	3	4 - 58 - 58	
13. How many members use the latrine	3	10 - 59 - 60	
14. Major source of drinking water	3	11 - 61 - 62	
15. Arrangement of garbage disposal	3	12 - 63 - 63	
16. Primary source of energy for cooking	3	13 - 64 - 65	
17. How often rubbin outside of communicable disease in the community	3	14 - 66 - 66	
18. Amount of medical insurance premium (Rs.)	3	11 - 67 - 74	
19. Household usual consumer expenditure (Rs.)	3	8 - 75 - 82	
20. Blank		44 - 83 - 83	
21. NSS	3	127 - 129	

You can just have a check on the layout file.

(Refer Slide Time: 31:45)

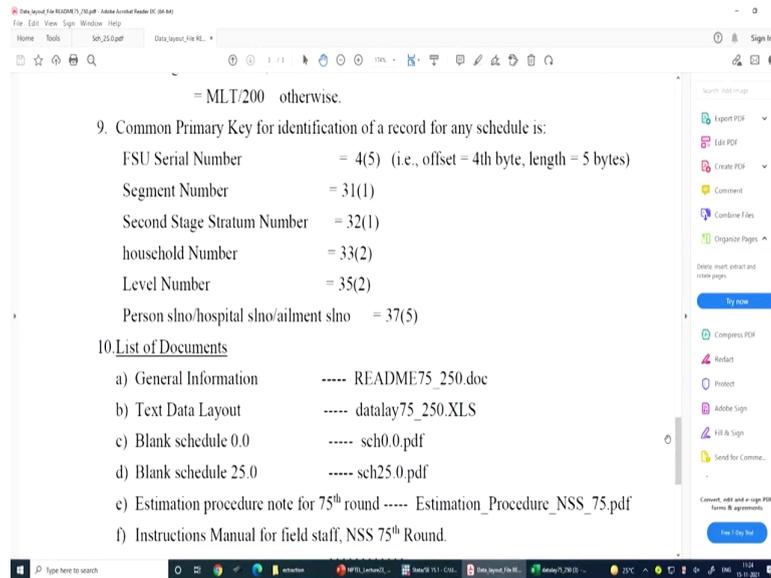
Item	Schedule reference	Length	Byte position	Remarks
1 Common ID		14	1 - 14	Auto-Appended
2 Level		2	35 - 36	"12" Generated
3 Film		1	37 - 41	"0000" Generated
4 Household size	3	1	2 - 42 - 43	
5 Whether SHD paid share fee	3	2	1 - 44 - 44	
6 SHD expenses for any non SHD				
7 NIC 2008 for age code	3	3	3 - 45 - 49	
8 NIC 2008 share age code	3	4	3 - 50 - 52	
9 Household type	3	2	1 - 53 - 53	
10 Religion	3	6	1 - 54 - 54	
11 Social group	3	7	1 - 55 - 55	
12 Type of tenure usually used	3	8	2 - 56 - 57	
13 Access to water	3	9	1 - 58 - 58	
14 How many members use the latrine	3	10	2 - 59 - 60	
15 Major source of drinking water	3	11	2 - 61 - 62	
16 Arrangement of garbage disposal	3	12	3 - 63 - 63	
17 Primary source of energy for cooking	3	13	2 - 64 - 65	
18 How often collect rubbish of community	3	14	1 - 66 - 66	
19 Amount of medical insurance premium (Rs.)	3	15	8 - 67 - 74	
20 Household usual consumer expenditure (Rs.)	3	8	3 - 75 - 82	
21 Wage	3	44	82 - 124	
22 SSC	3	127	129	
23 SSC	3	130	132	
24 MSLT	3	133	135	

This is level number 2 is written. Now it is household size then this is their NIC and National Industrial Classification 2 as per the 2008 and National Classification Occupation 2004 about their occupation. Household type, religion, social groups all such information with their byte position byte positions are also given.

So, byte positions and common id is also defined by default, but we are going to define the common id as per the instruction given in the layout file. So, common id at their byte position they have already gives 1 to 34 that is a combined common id, but we will define our common id the unique id which I have already told to you. And all the rest of the variables and the byte positions are already defined.

Now let us move to understand their common idea as well. So, these are the different files you have downloaded and you have checked their number of observations.

(Refer Slide Time: 32:51)



Here, are the common id related information in point number 9, common primary id or key for identification of a record is given. So, the common id as for the information for household level information, so this is household number is given. Then person number and hospital number or individual related number are given in here.

At the last line till this FSU segment number, second stage stratum, household number these four are important in this particular block. So, we will include 4 to 5 position, 4 and 5 means starting from 4 till 5th another 5th position. So; that means, 4, 5, 6, 7, 8 till 8 position you have to collect the information from the layout file.

Like it is already given FSU 4 to 8 you can just have a check, 4 to 8 here 4 till 8 in fact is your FSU serial number. This is what we need we have already done it in our extraction, I am not going to guide you further; how these byte positions are set we have already guided in our previous lecture.

Today after doing so, we know that these are the common id information. So, we have to extract this, and I take combined id for extraction. So, we have already done it based on that we have taken an FSU, SSS. This three and household, household number these four information are very essential.

So, that could also be combined and to be develop a household id HHID we can develop it I think in the PPT we have also guided you how you can combine this id.

(Refer Slide Time: 34:48)

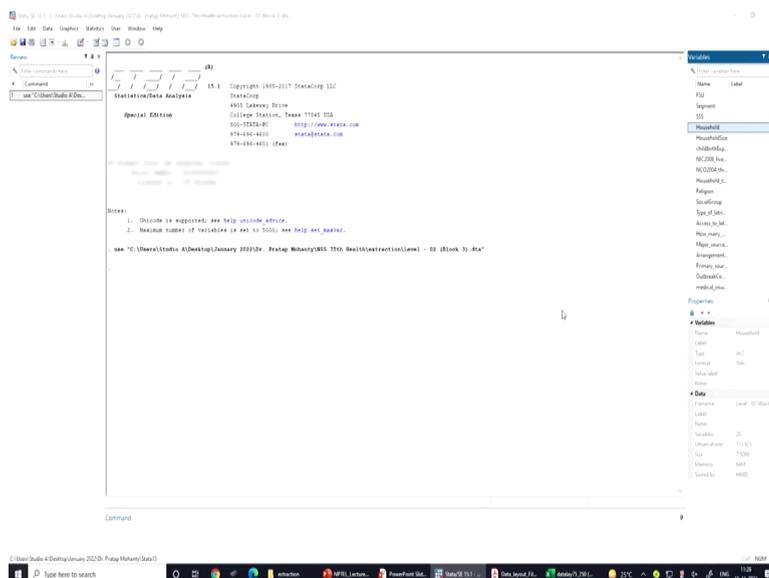
### Unique Id

- ❑ **Household:** HHID = FSU Segment SSS Household (Block 1\_2 and 3)
  - Command : `gen HHID = FSU + Segment + SSS + Household`
- ❑ **Individual Characteristics :** PID = FSU Segment SSS Household + Person Serial No.(block 4 and 5)
  - Command : `gen PID = FSU + Segment + SSS + Household + Person Serial No`
- ❑ **Ailment Persons:** Person Id = FSU Segment SSS Household Srl No of Member (block 6 and 7)
  - Command : `gen Person Id = FSU + Segment + SSS + Household + Srl No of Member`

23

So, that could also be combined like this. So, household id how HHID define like FSU segment, SSS household, they are for block number 1. 2 and 3. So, that is basically blocks number or level number 2 we in fact we are trying to deal with the household characteristics. So, this is FSU segment SSS and household this is what we have done it and defined the Id accordingly.

(Refer Slide Time: 35:16)



So, here on the window you have seen the HHID. So, now, I am just going to show all those things together that is browse HHID i.e., br HHID.

Then other three FSU then segment then SSS and then household. So, I am just going to compare how these are defined.

(Refer Slide Time: 35:46)

HHID	FSU	Segment	SSS	Household
0000000000	00000	1	0	00
0000000001	00000	1	0	00
0000000002	00000	1	0	00
0000000003	00000	1	0	00
0000000004	00000	1	0	00
0000000005	00000	1	0	00
0000000006	00000	1	0	00
0000000007	00000	1	0	00
0000000008	00000	1	0	00
0000000009	00000	1	0	00
0000000010	00000	1	0	00
0000000011	00000	1	0	00
0000000012	00000	1	0	00
0000000013	00000	1	0	00
0000000014	00000	1	0	00
0000000015	00000	1	0	00
0000000016	00000	1	0	00
0000000017	00000	1	0	00
0000000018	00000	1	0	00
0000000019	00000	1	0	00
0000000020	00000	1	0	00
0000000021	00000	1	0	00
0000000022	00000	1	0	00
0000000023	00000	1	0	00
0000000024	00000	1	0	00
0000000025	00000	1	0	00
0000000026	00000	1	0	00
0000000027	00000	1	0	00
0000000028	00000	1	0	00
0000000029	00000	1	0	00
0000000030	00000	1	0	00
0000000031	00000	1	0	00
0000000032	00000	1	0	00
0000000033	00000	1	0	00
0000000034	00000	1	0	00
0000000035	00000	1	0	00
0000000036	00000	1	0	00
0000000037	00000	1	0	00
0000000038	00000	1	0	00
0000000039	00000	1	0	00
0000000040	00000	1	0	00

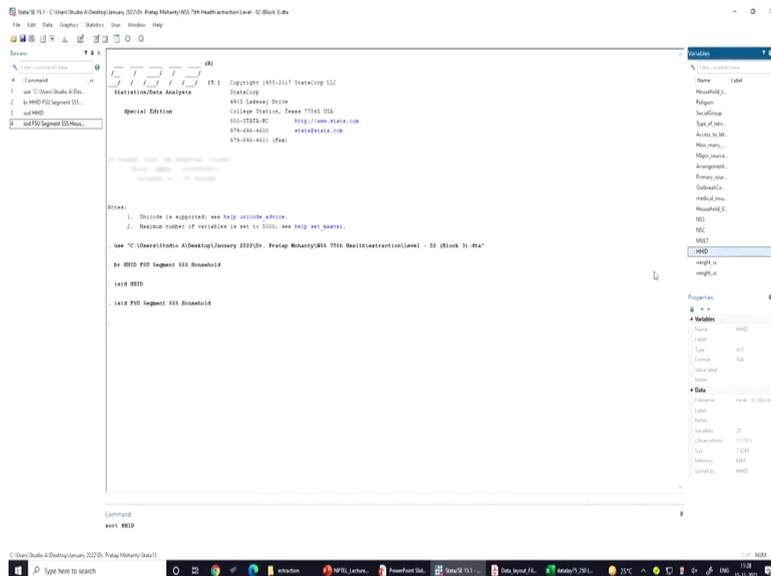
So, look at just these four carefully. So, this HHID is actually taken together FSU segment SSS and household if we combine all those things, then that is defining our HHID. How these are actually added together these are added together because these are in string entries.

Since these valuable are in string those could be added, but not by numerical value, but these are simply kept together. So, that is all about the id defined.



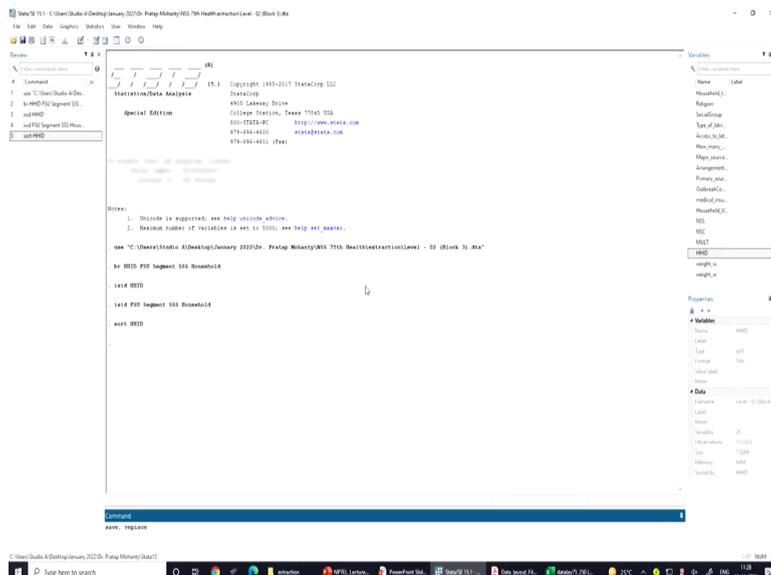
unique id by taking these three together isid, FSU, Segment, SSS and household these four together FSU, segment, SSS and household.

(Refer Slide Time: 37:19)



So, enter this is also not showing any error. So, that means, this is perfectly identify and what is our first step then we need to sort this id variable that is HHID. Let us sort this HHID which you have already defined.

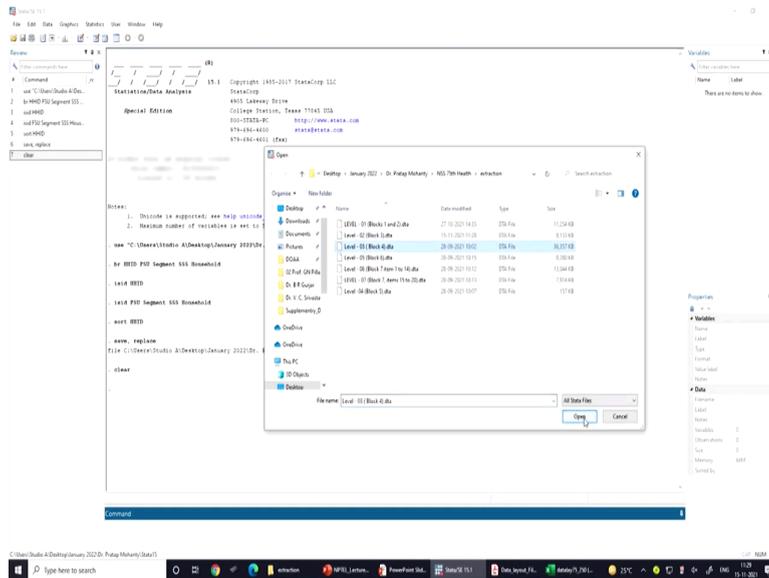
(Refer Slide Time: 37:38)



Then we need to save it we have to also check our other using file.

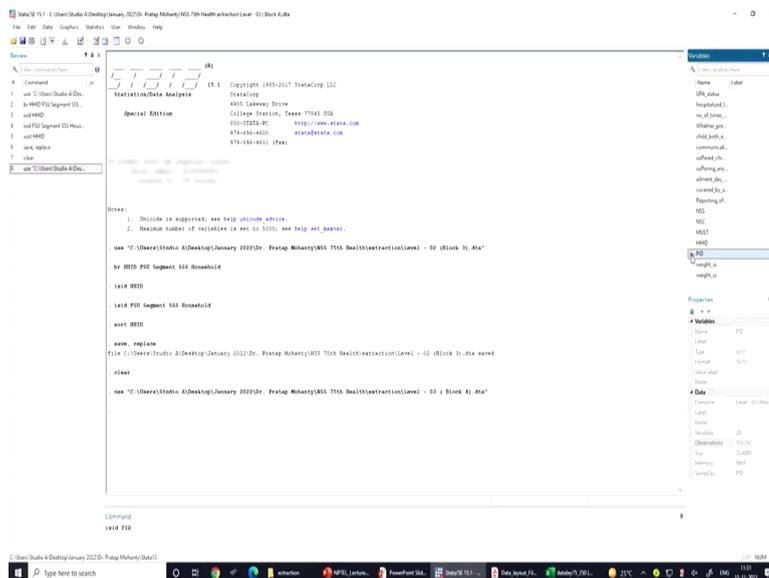


(Refer Slide Time: 38:05)



Now, we are going to open, let it be level number 3.

(Refer Slide Time: 38:10)



So, now we have opened similarly for the level number 3, that was extracted on the previous day, regarding that we have guided. Similarly, we have defined the HHID, but HHID is now little different. Because what is these 3 these 3 is in fact our individual id just look at age of the person, age of the household cannot be asked. There is no age for the household, it should be age of the person.

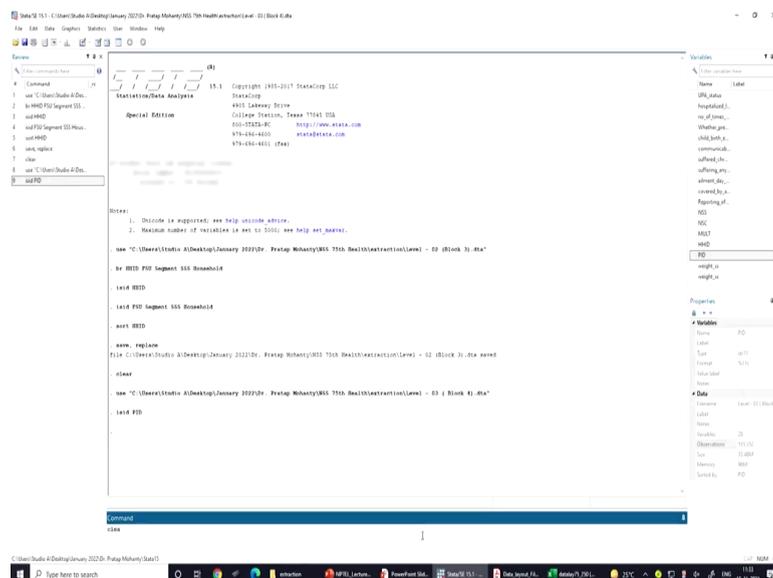
Similarly, gender, what is the gender of a person. So, this is an individual file. So, what do you mean by individual file. You just check how many observations are there, observations in the level number 3 is 5,55,352. You need to just cross check once again with the observations given displayed on the Stata that is 5,55,352 that is then perfectly fine.

And next one is when we say that this is an individual file what is it is unique id variable is, if it is in individual file then person serial number has to be included as well along with the household number. So, how you have defined we have also guided in this particular PPT page.

And on this particular page PID that is personal individual Id variable. Personal Id variables is FSU, then SSS then person serial number. So, all other information along with the person serial number is also important. So, that is PID, FSU, then segment, then likewise the previous all four are carrying. So, this is what we have done it and I am going to show it again.

So, now we need to check that we have made this PID here, not HHID. HHID was already there in the previous one. So, for us it is PID, PID we can just check whether it is uniquely identified or not: isid PID.

(Refer Slide Time: 40:26)



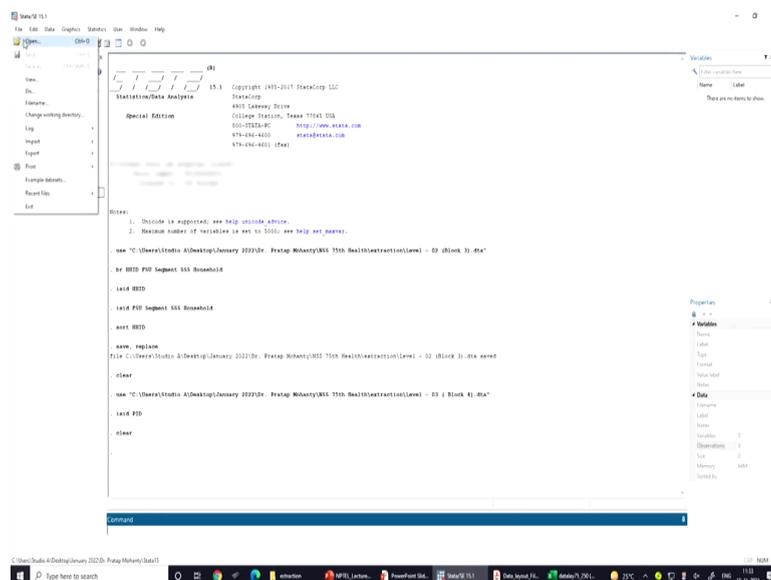
So, PID is now not throwing any error, it is uniquely identified. Two aspects to be remembered at this point, if you are keeping this as our master file, if you wanted to add

another file to this particular file like if the individual file is added with the household file, in that case this PID has to be uniquely identified and in another case PID may not be uniquely identified, because which is a different file. But if you are merging from household to individual then what happens, where the cases are many or ids are many and the individual file it is many.

So, the command is going to be different now, if this is your master file then your command is what? Your command is master to household if you are doing it household is having less number of cases. So, it is going to be m is to 1. So, m is your individual file having more observation and it going to be merge with the 1 that is your household.

So, your household file has to be uniquely identified. Now for your simplicity the same HHID that was carried, this is not important much, but we will clear this file and go to the master file once again.

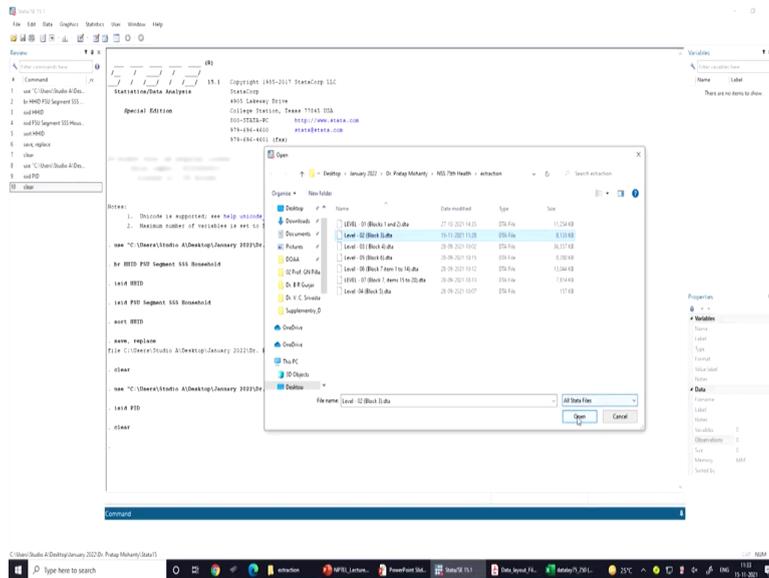
(Refer Slide Time: 42:08)



```
use "C:\Users\Studio\Desktop\January 2022\Dr. Prateep Mohanty\BMS 75th HealthRestrictionLevel - 02 (Block 3).dta"
merge 1:1 HHID PID logmerge 1:1 Household
clear
use "C:\Users\Studio\Desktop\January 2022\Dr. Prateep Mohanty\BMS 75th HealthRestrictionLevel - 03 (Block 4).dta"
merge 1:1 PID
clear
```

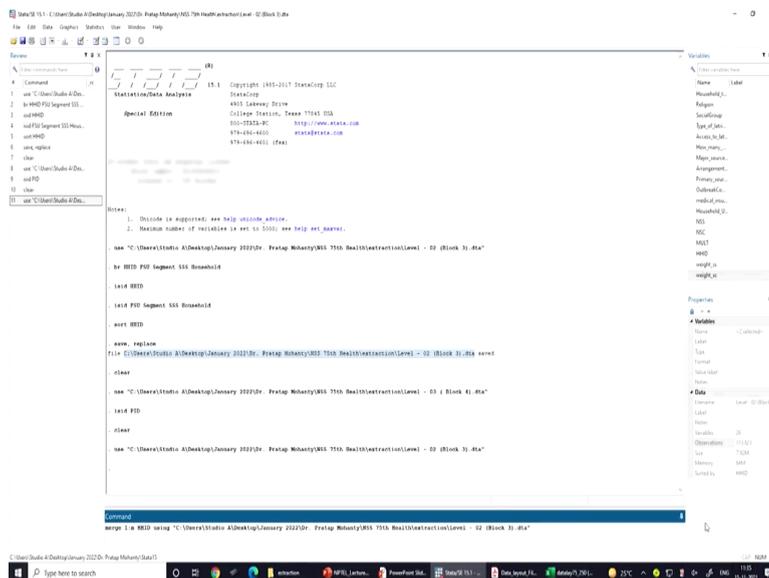
So, master file that is the household file we will go once again. So, let us open the household file again on the screen.

(Refer Slide Time: 42:16)



So, now we have opened that block number level number 2.

(Refer Slide Time: 42:20)

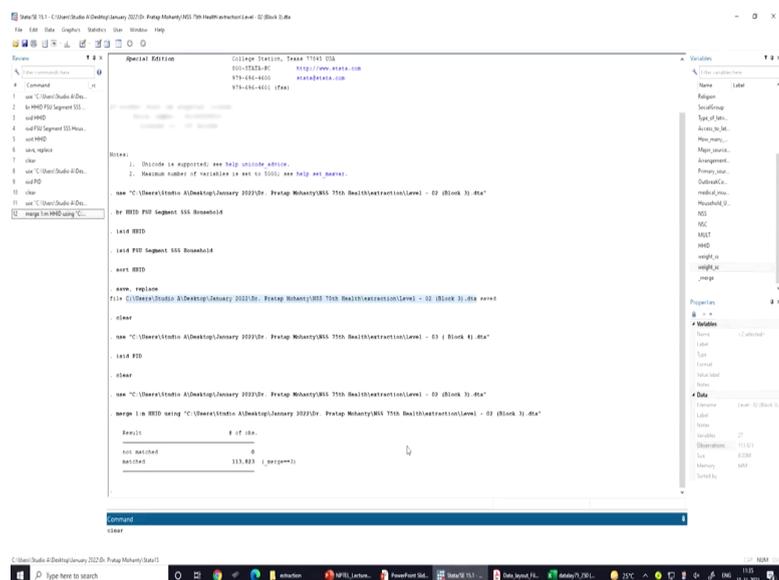


It is on your screen. Now this is open and we know that our unique Id file is uniquely identified that is HHID is uniquely identified and that was also sorted. So, now, we are going to merge these two files. So, what command we are supposed to give it has to be since our master file is 1 and our that is household and your merging using file is many that is individual.

So, it has to be 1 is to many and which one should be uniquely identified that is your master file since it is 1. So, it is merge, merge 1 is to many, 1 is to m, then it is using file using a path name. Then, 1 is to m then we have to give our the specific unique id that is HHID we have already find, so then using file.

So, using file is your individual file and it has to be mentioned with the path name with the extension file as well. So, here it is with its path name it is in inverted comma. Now you can just check number of observations are initially 1,13,823, after merging it will be increased as with the individual file that was something around 5,55,000 just enter you will get the merging file.

(Refer Slide Time: 43:56)



```
data _null_;
  set 'C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 02 (Block 3).dta';
run;

data _null_;
  set 'C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 03 (Block 4).dta';
run;

merge 1 in BBTD using 'C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 02 (Block 3).dta'
      by HHID PID Segment SSS Bazaarhold;

list BBTD;

list PD Segment SSS Bazaarhold;

run;

proc print;
  title 'Merge 1 in BBTD using 'C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 02 (Block 3).dta'';
  var HHID;
  var not_matched;
  var matched;
run;
```

Log

```

1. Divide in appropriate; see help on PROC MERGE.
2. Maximum number of variables is set to 2500; see help on MERGE.

*** C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 02 (Block 3).dta ***
BY BBTD PID Segment SSS Bazaarhold

LIST BBTD

LIST PD Segment SSS Bazaarhold

*** C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 03 (Block 4).dta ***
LIST PD

*** C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 02 (Block 3).dta ***
MERGE 1 in BBTD using 'C:\Users\Studio\OneDrive\January 2022\Dr. Prateep Mohanty\MS 7156 HealthRestrictionLevel - 02 (Block 3).dta'
      BY HHID;
      NOT MATCHED          0
      MATCHED              113,823 (_merge=1)

```

Variables

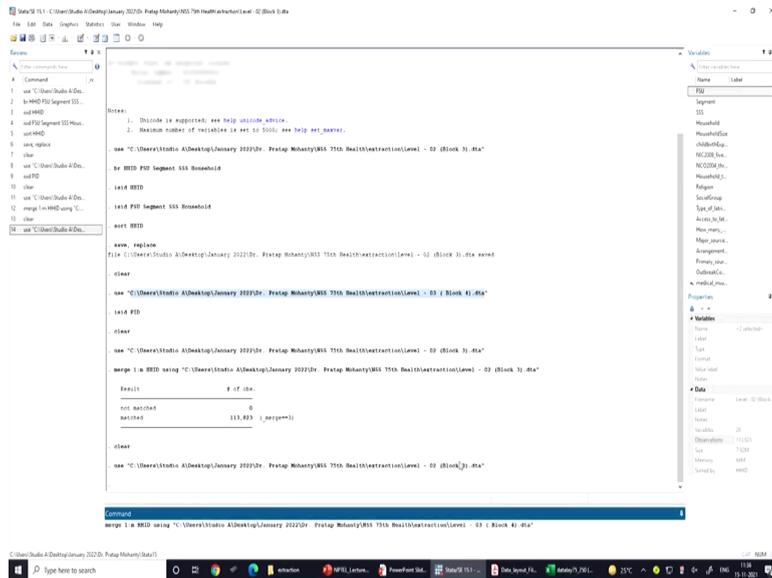
Name	Label
HHID	
PID	
Segment	
SSS	
Bazaarhold	
not_matched	
matched	
_merge	

Command

```
clear
```

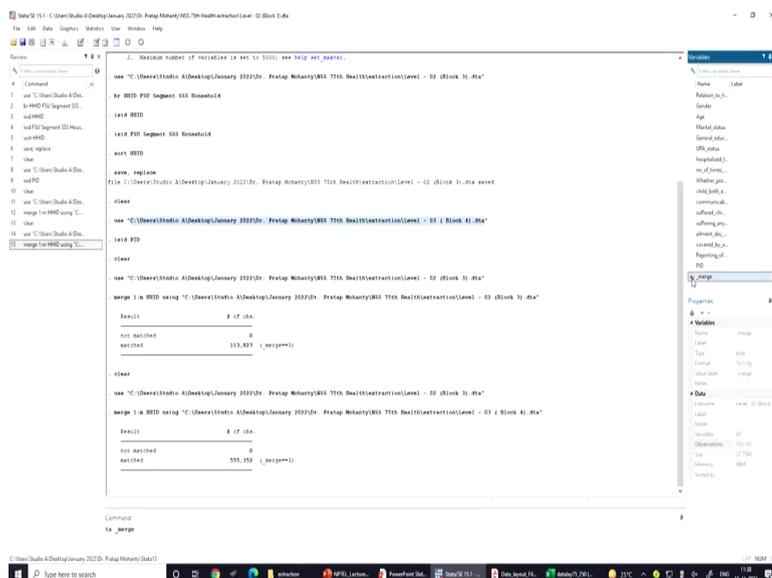
Now, at this moment we have got how many matched exactly, but the number of observations that are merged with what exactly merged. We need to clear once again and operate it.

(Refer Slide Time: 44:32)



We are not going to save it, we will clear it and merge it once again. So, this is 2 open on your screen. So, 2 is perfectly fine, then you need to check that merge file is repeated or not, dot underscore merge whether it is available on the variable list or not. So, HHID it has to be correctly spelled with its capital letter, then using it has to be which one you ask we wanted to use the third one not the second one. So, now inverted and within inverted comma then enter.

(Refer Slide Time: 45:29)



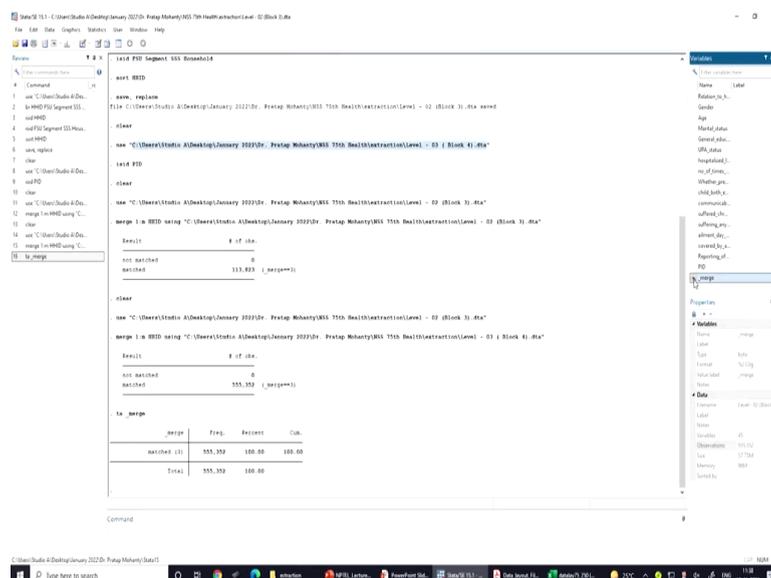
Now you can see that initially we did one small mistake that with we added the same file. We merge the same household file that's why the number of cases were not increased, but now you can see the number of cases increase to 5,55,352.

Now another interesting aspect is that initially we carry the file with household characteristics, but now you have got characteristics about the individual information, gender, age, marital status etc. on the same file. So, now, the number of matched cases are 5,55,352. So, the previous one since it was an error, so you need to avoid the last one. That we have already guided to you.

Similarly, we have all other records we have specified in our PPT that will be quite useful for you to go through this is the one you can do. There are other files also extracted file like, one now after extract after merging of these two files the new file is coming up with one important variable that is automatically added called merge, that merge is created underscore merge.

You can also check the merged information with its frequencies like ta you can type on the screen and just click on this merge file.

(Refer Slide Time: 47:03)



You will get the information like how many observations are there, how many are matched. So now, one important suggestion to you is that if again you want to merge another file

maybe block number 5, maybe block number 6 you need to delete or rename this particular underscore merge file.

Otherwise, Stata is not going to read your command. That is one suggestion must be noted down very carefully. I think these are important guide directions we have already given and all other suggestions that are important, I have already given in between.

But still if anything is missing, I can guide you very quickly, I am just going to check the previous one that whether I have guided or not. Like what is important again once you have merged a file, your new file is no longer a household file it is an individual file now because individual information have been already added.

Now further if you wanted to merge another file, your first file if you are open this merged file this is going to carry as your individual file not your household file. So, I am just checking whether I missed anything that was discussed in between I think I would almost all information I have emphasized through the practical session. And the last slide we have kept for your guidance is to create some unique id.

Like there are for ailment persons file you need to add the serial number of the member; who has reported as ailment, rest are same. So, these are all guidance for your merging I hope you can be able to get your fine-tuned data, but even if you have done your merging still it requires fine tuning data recoding, some developing composite variable before you going to have your analysis.

These are all details we are going to guide you in the next week, but this week is meant for clarifying about the data set for merging and appending.

I hope you have enjoyed this lecture and the similar kind of guidance we will give it on the next lecture for sure. With this I am going to stop here we will continue on the next class.

Thank you.