### 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)



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

(Refer Slide Time: 04:23)



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.

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)



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 countered 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)

								/aria	bles			
		ENTID	CentreCode-d	Vill_Blk_Slno	Round	Schedule	Sample	Sector	_**C\$0Z	State_Region	State	_* ^
	1	6950911802	738	69509	73	234	1	1	roral	091	0.9	
	2	\$3\$3911101	730	83839	73	234	4	1	urban	031	03	
	3	\$116511403	738	51165	73	234	1	2	urban	212	21	
	4	6770351101	733	67783	73	234	1	1	rorel	245	24	
		6760911403	73 <b>A</b>	67.689	73	234	1	1	rurel	241	2.4	
	6	8777411403	7 337	07774	73	234	1	1	urban	272	37	MA
	- 1	6916310702	735	69163	73	234			uzban	091	09	
		0001511404	737	60615	73	234			FUTAL	151	15	
	- 10	5313111503	790	50171	74	104			urben	041	0.0	
	T	5963011404	735	59630	73	234	÷	2	urban	092	09	
	12	5603711601	730	56037	73	234	1	1	rural	321	32	
	13	8193510201	738	01935	73	234	1	2	urban	193	19	
Observation	14	9076711402	737	50767	73	234	1	1	rural	142	14	
Observation	15	5530711502	738	55307	73	234	1	1	rural	111	33	
	16	5515310201	738	95193	73	234	1	1	rural	341	34	
	17	\$355011604	730	53550	73	234	1	2	urban	061	06	
	18	5023610001	73F	60236	73	234	1	1	uzban	194	19	
	19	6966711501	732	69667	73	234	1	1	rural	091	0.9	
	20	\$733211401	733	57332	73	234	1	2	uzban	242	24	
	21	6736210603	73 <b>A</b>	67362	73	234	1	2	urban	241	2.4	
	23	€853521501	780	68939	73	234	1	1	rural	111	11	
	23	6158920101	735	61989	73	234	1	1	rural	192	19	
	24	6379011602	730	63750	73	234	1	1	rural	032	03	
	26	0969411302	738	89694	78	234	1	2	urban	095	09	*
	<											,
🎯 swayalli 🛞												

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)



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)



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)



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 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)



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)



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)



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)



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)



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 ideal 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 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)



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)



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		
0	Swayam 🧕	22	

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

(Refer Slide Time: 27:37)



So, here we are going to open Stata here. See in the data set we are just going to open first some important information which are required for merging.

(Refer Slide Time: 27:49)

iw 14:	κ 🗌							<ul> <li>Variables</li> </ul>
Liter commands have								A Liter variables here
Command re	1-1-	8 1 1	Manage State15				- U x	Name Label
Date are no dates to show	Statistics/	He Home Share View	App Too's				- O	There are no itered to show
		+ · · + 📒 + January 2022 +	Dr. Pratap Mohanty + Stata15				✓ δ	
	Special							
		Quick access	harte	Date modified	type	204		
		E Desitos d	edu -	27-10-2021 11-42	Filefolder			
			deci	27-10-2021 11:44	F da foldar			
			ginap	27-10-2021 11-44	Filefolder			
			- utitier	27-12-2021 11:44	Filefolder			
			exto.dta	07 05-2018 00-16	USA Pile	718		
		DOAA Y		12-03-2020 18:98	035734	12.83		
	Notes:	C2 Prof. GN Pillai	C) common completion	10.00.0000.00.00	Diff for	7.0		
	1. Dos	Dr. B.R. Gurjer			00.54	110		
		Dr. Pratap Mohanty	ana ana	13-01-3020 19-47	Wellie 29 writes	(5.42)		
	- F	Dr. V. C. Srivastav	instate 190	07 06 2018 00 16	150 File	10		
		<ul> <li>Outbing</li> </ul>	II lay state	11-05-3020-00.79	Text Document	1.01		
		- Colored	mianodistricts-coordinates da	12-03-2020 18-96	014 File	41.13		
		OneDrive	mlangfistricts-database.dta	12-03-2020 18:56	OSA f de	1.43		
		INK	miantibutine coordinates.tha	12-03-2020 18:56	0%A.File	14.13		Properties
		10 Objects	milanopolice-database.dta	12-03-2020 18:56	D3h File	1.63		A
		Deritan	publics cresitals_test.do	10-05-2020 19-25	DO File	1.18		+ Variables
		2 Description	putdocr/regtable_test.do	15-05-3030 18:33	DO File	1.63		Name
		Documents	replicate_lalonde.do	18-05-2020 17:06	00 File	1.13		Label
		Downloads	🖗 stata	07-06-2018-00-16	Chrome HTM, Do.,	1.63		1 <sub>000</sub>
		Maric	STATALIC	12-06-2018-01-99	LIC File	1.0		Format
		👔 Potuns	StatuSI-64	21-02-2020 14:54	Application	45,229 KB		Notes and
		Videos	StataSE-64_old	10-01-2020 22-45	Application	65,322.89		+ Data
		Local Oick (C.)	M trap2-userguide	12-03-2020 18:96	Adobe Acrobat D	COLOR		Liename
		USB Drive (D.)	us40-coordinates.dta	12-03-2020 18:56	O'SA Pile	191108		Label
		Local Disk (E)	- Us coordinates.ats	12-03-2020 18:98	00.14	213.68		Neter
		Local Disk (F)	C C MANNEAU		0.24 2.00	51.0		
		USB Drive (D.)						Memory 64M
		NSS 79th Health	·					Sorted by
		27 items 1 item selected 63.7 MB					12 K	
	Command							

# (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)

	x								<ul> <li>Variables</li> </ul>
er commandichara	0	📙   🕑 📙 🖬   Supplemently	Document					- 0 ×	A filter variables here
ommand _rc	1 1 1 1	File Home Share	Ver					- 0	Name Label
here are no items to show.	Statistics/	4	- 2022 - 0	Roman Madalanta - 1955 Talk Maddle - Care	dependent Decomposition			a A O Sand Contempto Day	There are no items to show
	Special		,		person jurico de la companya de la c			<ul> <li>G &gt; An or Approximity Sect.</li> </ul>	
		Ouisk scene	- 1	Name	Date modified	lype	Size		
		Desiton		🔒 Appendix-I	25-09-2021 11-27	Adobe Acrobat D.	101.48		
		Tardetak.		Data_layout_File README75,250	25-04-2021 11-26	Adobe Arrobat U.	.15 KB		
				Mature 15,230 (1)	27-10-2021 14-15	Monet Erel IT.	54 KB		
			- 1	A 505,00	25-09-2021 11:27	Adobe Acrobal D.,	334.430		
			- 1	B 200,250	25-09-2021 11:26	Adobe Acrobat D_	892.68		
		DOAA	1	~					
	Notes:	C2 Prof. GN Pillui							
	1. 0hs	Dr. 8 R Gurjar							
		Dr. Pratap Mohanty							
		Dr. V. C. Srivastav							
		<ul> <li>OneDrive</li> </ul>							
		<ul> <li>OneDrive</li> </ul>							
		💻 This PC							Properties
		30 Objects							8.11
		Desktop							<ul> <li>Variables</li> </ul>
		Pocumenta 6							Name
		& Downloads							Label
		h Masic							format
		E Febru							Value label
		Videos							Noter
		interest and a second s							* Data
		Locardex (c.)							
		<ul> <li>Use bive (D)</li> </ul>							
		Local Disk (E)							
		Local Disk (F)							Obsenations 0
		USB Drive (D.)							Sax 0
		NSS 75th Health	¥						Memory 64M
		Sitema 1 dam selected 891	a .					Ex.	
	Command								
	Contratio								

So, we open schedule.

(Refer Slide Time: 28:00)

1. 0 ~	🗇 🕘 1 / 16 🖡 🕙 \ominus 🕘 1185 - 🛗 🐺 🗒	1 4 3 1 0	<i>i</i> 2
RURAL     *       URBAN     *       So     [0] descriptive identifica       1. state'UT.:     2. district:       3. sub-district/tehsil/town	GOVERNMENT OF INDIA NATIONAL SAMPLE SURVEY OFFICE SOCIO-ECONOMIC SURVEY SEVENTY FIFTH ROUND: JULY 2017 - JUNE 2018 HEDULE 25.0: HOUSEHOLD SOCIAL CONSUMPTION: ion of sample household 5. hamlet name: 6. investigator unit no . block no 7. name of head of household.	CENTRAL * STATE *	Add Hogy     Add Hogy     Add Hogy     Add Hogy     Comprese Rof     Comprese Rof     Comprese Rof     Comprese Rof     Comprese Rof     Adder Sign     Adder Sign     Adder Sign     Send Lo Commo     Kore Tools

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

(Refer Slide Time: 28:08)

![](_page_22_Picture_1.jpeg)

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

(Refer Slide Time: 28:27)

Auto	Same 💷 😸 🗧	· (* 8 + +		nday 19,21	0(1) - Cong	utbilly Mod	e •		P	Search										- 4	ernesh Arber	unter (		-	ø x
file	Home Insert	Drave Page Layout	Formulas	Data	Review	riev He	lp .																d 9	kare 🖓 i	Comments
	λ Cut Di Comi a	Times New Roman - 11	- A' A' = 3		8- A	Avap Text	Gen	eal			<b>R</b>	Normal_She	Normal	Bad		Good			<b>B</b> (	8,	∑ AutoSum ∎ Fill •	· * 27	Ω,	0	
	of Format Painter	81⊻-⊞-≼	• <u>4</u> · ≣ 3	53	3 2 🗒	Merge & Cer	ter - 🐻	% )	14	Conditional For Formatting * Ta	vratas   ible -	Neutral	Calculation	Check(	લા	Explorato	Y	liset	Delete Fo	rmat i	Clear +	Sort & Filter	2 Find &		
	Cipboard 5	fert	6		Alignment		5	Number	5				Shies						Cells			Editing		Analysis	
289		V 14																							
11		8	C D	ε	F	GH	1	1	-	K		L	М	N	1	0 F		0	R	5	T	U	V	W	X
1		Text L	Para Layour																						
2		NSS 75th Kound	(Jelý 2017-June 2	2018)																					
4				Total	io of levels	= 13																			
5	Sch. 25.0 : LE	VEL - 01(Blocks 1 and	2)	Recor	d Length =	142+1																			
6									_																
7 8	ι	Item	Schedule refe	resce	Leagth	Byte posi	lion	Remarks																	
8 0	<u> </u>		Block Item	Col																					
2	1 Centre, Round				- 1	11.	1	Generated																	
20	2 PSU Serui No.		·																						
	5 Kouse		·	÷	- 1	· · ·		Cenerated																	
12	<ul> <li>Scorela</li> </ul>		· · · · ·				10 c	of Generation																	
4	6 Sector				1	15	15 Ceneral	of free Sch.																	
15	7 NSS-Region				·	16	13 General	ed from Sch	0.0																
15	8 District				2	19 -	20 General	ed from Sch	0.0																
17	9 Stratum				2	21 -	22 General	ed from Sch	0.0																
18	0 Sub-stratum				2	23 -	24 General	ed from Sch	0.0																
19	11 Sub-Roard				i 1	25 -	25 General	ted from Sch	0.0																
20	12 Sub-sample				1	26 -	26 General	ed from Sch	0.0																
21	13 FOD-Sub-Region				4	27 -	30 Genera	ed from Sch	0.0																
22	14 Hanlet group' Sul	b-black na	1	4	1	31 -	31																		
23	15 Second-stage-stra	han no.	1	5		32 -	32																		
24	16 Sample hhld. No.		1 1	6	2	33 -	34 **(	ommon-ID*	۰.																
25	17 Level				2	35 -	36 10	0° Ceserated																	
26	18 Film					37 -	41 '00	000, Eurusti	st .																
27	19 Sl.no. of informan	t(as in col. 1, block 4)		1	. 1	42 -	43																		
18	30 Kespinse Code		· 2 3	÷		** *	**																		
2	11 Salvey Cone	Combineda	- 1 1	2		*) -	42																		
11	13 Employee code	Castany Cool	2 1(4)(6	ĩ.	i 1	47	50																		
22	14 Employee code		2 1(4)(6			9	u u																		
33	25 Employee code		2 1(b)(#	5	4	55 .	58																		
34	36 Date of Survey		2 3(0)		6	59 -	64 T	D MM YT'																	
35	27 Date of Despatch		2 2(W)	1.1	i 6	65 -	70 T	D MM YT"																	
36	28 Time to canvass (	minutes)	2 4	1	3	71 -	73																		
37	29 Ne. ef investigato	rs (FI ASO) in the team	2 5		1	74 -	74																		
38	00 Remarks in block	12/13	2 6(i)	1	1	75 -	75																		
	layout75	25.0 (+)																							
Feedy																					8		四		+ 12
	0.5		0	Hi						MIT Laborati	10	0	em 🖪				ma				A 11			na IV	22
1	> type here to	search	0	H	<b>a</b> .	e	atuo			in minerares		and a pro-	- C	númeta - s		a analy	(con D) -		0.0		- L		- 0°	15-11	2021

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)

AutoSee 🚥 🗟 🗉 - 😳 8 = 1		J <sup>D</sup> Statch		Remesh Anterandum 😣 🖽 – 🖉 🗙
File Home Insett Draw Page Layout Form	nulas Data Review View Help			d Share 🖓 Comments
A 100				Σ Autolum - Arra O I Tast
Innes New Roman - 11 - A A	General General General	Normal_She Normal	Bad Good . See 25 1	1 2V ) [2]
- of format Painter B I U - 🕀 - 🧕 - 🧾	🛛 🔁 🧧 🖬 I estruction			X Q Clear * Filter * Select * Data
Clipboard fs Fort	de Home Share View			toting Analysis A
m	and a state of the	with a seturation	. A. O. Sandardantes	
145 · · · · · · · · · · · · · · · · · · ·	· · · · · · · ·		• • • • • • • • • • • • • • • • • • •	
4 A B	A Name	Date modified Type 1	Sce	TUVWX.
1 I CAT Data	Desites	3 dta 27-10-2021 14:35 01% File	11,254.K8	
2 355 /50 Kould/July	Level - 02 (Block I).dts	27. 15. 3104 5.8-30 PHS Fda	8,133 KB	
4	Level - 03 (Block 4) dta	How do you want to open this file?	36,337.08	
5 Sch. 25.0 : LEVEL - 01(Blocks 1 and 2)	Entern f Level - 10 (Block Edds		12 MARY	
6	DOAA e Dittyl - 07 (Beck 7 ee	Look for an app in the Microsoft Store	2,04:00	
7 srt. Item 1	D Prof. GN Priloi		157.08	
9   Centre Roard	Dr. R.R. Gurier	More apps 4		
10 2 FSU Serial No.	D. Y. C. Sivester			
11 3 Round	Supplementry Documents	Always use this and to open dta files		
12 4 Schedule				
13 Sample	OneDrive	<u>b</u>		
15 7 NSS-Region	CheDrive	.40x		
16 8 District	Dia K.			
17 9 Stratum	30 Objects			
18 10 Sub-straten	Desktop			
17 11 Sub-Kound 20 12 Sub-connele	Documents			
21 13 FOD-Sub-Regim	& Downloads			
22 14 Hamlet group' Sub-block no.	h Mais			
23 15 Second-stage-stratum no.	E Potern			
24 16 Sample Mid. No.	Videos			
26 13 Film	Local Disk (C)			
27 19 Sino. of informant(as in col. 1, block 4)	USB Drive (Dr)			
28 20 Response Code	Local Disk (E.)			
29 21 Savey Code	🕳 Local Disk (F.)			
32 22 Substitution Code: Casualty code 33 21 Employee code	100.001			
32 24 Employee code	Set onve (o)			· · · · · · ·
33 25 Employee code	NGS (20) PRARY			
34 26 Date of Sorvey	e nov and the second the second at the secon		E	
35 27 Date of Despatch	2 2(iv) 4 6 65 - 70 "DD MM Y"	·		
36 25 Time to canvass (minutes) 37 29 Na. of investigators (EU ASO) in the term	2 4 3 3 71 - 73			
38 30 Remarks in block 12-13	2 6(0 3 1 75 - 75			
i i izyout75,25.0 (+)				
Testy				■ □ □ - + · 125.
	0 H 6 d 0 I mutu	1 MIR (1997)	Data barren film M	A C D A 4 6 MM 1120
ype nere to search	на 🤍 — 🚺 налодом	Manufactures	and appendix and a manager 2000 - 2000	15-11-2021

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

(Refer Slide Time: 29:09)

![](_page_23_Figure_4.jpeg)

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

(Refer Slide Time: 29:11)

![](_page_24_Figure_1.jpeg)

(Refer Slide Time: 29:16)

14 Data Combine Data	Atopijanuary 2022/Dr. Fratap Mohanty/NSS 79th Healthi art	action/J,evel - 02 (Block 3).dta	- 8 )
S   ∃ - 1 15 - 13	1 in 11 0 0		
ev T a	x		- Variables T 0
Liber commands here	a (a)		A filter vanables have
Campand I at			Notes Label
une "Cillken/Stuffe #/Des.		Copyright 1993-2017 StateCorp LLC	- F9
New Surgering and the second	statistics/sets Assignie	4905 Lakeway Drive	Gunda
	Special Edition	College Station, Texas 77045 USA	222
		800-STATA-PC http://www.stata.com	Household
		979-696-4600 stats@stats.com	HeuseholdSize
		979-696-4601 (fax)	challiothExp
			NEC2000_5vm
	the state of the second		NCO2004_bhr
	second or of the second		Household_t
			Religion
			SocialGroup
	Notes:		Type_of_latri
	1. Uhicode is supported) see	telp uticode advice.	Access_to_lat
	2. Nether supply of verseye	a set to poor, see help set manent.	How,many,
	. use "C:\Dsers\Studio A\Desktop\Jan	sary 2022\Dr. Fratap Mohasty\MSS 75th Bealth\extraction\Level - 02 (Block 3).dta"	Major_source
			Anangement
			Primary_sour
			Outpreak.c
			medical_mov
			Properties 9
			A
			* Variables
			1
			1997 Lionat
			Viba label
			Noter
			* Data
			Liename Level - G/ (Dpc)
			Label
			Fister
			Variables 26
			Sax 7.92M
			function in the second
	Command		- 4
s/Studio A/Desktog/January 2022	Dr. Pratap Mohanty/Stata15		C4P_NUM

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)

ne Tools Sch. 25.0.pdf 0	ata Jayout, Fie RL ×		🕐 🏨 Sign In
☆ ⊕ ϴ Ϙ	0 🕘 <u>1</u> /3 🖡 🕙 🖸		<i>C</i> _ 🖂 🤇
	File names	No. of Records	Search 'Add Image'
	R75250L01.TXT	113823	💽 Export PDF 🗸 🗸
	R75250L02.TXT	113823	Edit PDF
	R75250L03.TXT	555352	Create PDF 🗸
	R75250L04.TXT	2537	Comment
	R75250L05.TXT	93925	Comone rives
	R75250L06.TXT	93925	Delete, insert, extract and institute manes.
	R75250L07.TXT	93925	Try now
	R75250L08.TXT	43240	<ul> <li>Compress PDF</li> </ul>
	R75250L09.TXT	43240	🔏 Redact
	R75250L10.TXT	43240	Protect
	R75250L11.TXT	42762	Adobe Sign
	R75250L12.TXT	70258	🔓 Send for Comme
	R75250L13.TXT	32257	- Convert with and assiss PDF
	Total	1342307	forms & agreements
Pagord la	ath for data is 142		Free 7-Gity Inal

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)

    1  100	nulication of sample nousenoio		~						
									Search 'Add Image'
item	item		code		item	item	code	0	Export PDF 🗸 🗸
10.	of up of sounds FSU				10.	counts have dealed much as			Edit PDF
1.	sti. no. of sample FSU		1		0.	sample household number			Create PDF V
2.	round number	7		5	7.	serial number of informant (as in column 1 of block 4)			Comment
3.	schedule number	2	5	0	8.	response code			💱 Combine Files
4.	sample hg/sb number				9.	survey code			扪 Organize Pages 🔺
5	second-stage stratum number				10	reason for substitution of			Delete, insert, extract and rotate pages.
5.	second-stage straum munoer				10.	original household (code)			Tiy naw
CODE	S FOR BLOCK 1							,	Company PDE
item 8:	response code: informant: co-opera	tive and	cava	ble -1. c	o-overat	tive but not capable -2, busy -3, r	eluctant -4		A Reduct
	others -9.								A RECARD
item 9:	survey code: original -1, substitute	2, casua	ilty -3						Protect
item 10	): reason for substitution of original	househ	old: in	forman	t busy -1	, members away from home -2, i	informant		Adobe Sign
	non-cooperative -3, others -9.								🔏 Fill & Sign
									🔓 Send for Comme
* tick n	nark ( V) may be put in the appropriat	te place							
									Convert, edit and e-sign P

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

## (Refer Slide Time: 30:20)

2] particulars of f 3l. item (1) (a) (i) name (bloo (ii) code	(2) 2 / * k (2) teld operations       (2)       (2)       (k letters)	⊙ ⊙ Fiel Junio	d Inves or Stati	stigato istical ISO) (3)	T (FI) / Officer	L &	D D I Officer istical O	(FO) (ficer	Senior (SSO)	-	Î	Search 'Add Image
2] particulars of f sl. item (1) (a) (i) name (blow (ii) code	(2) k letters)	Fiel	d Inves or Stati (J	stigato istical ISO) (3)	r (FI) / Officer	Fiel Sta	l Officer istical O	(FO) (ficer	Senior (SSO)	-	Î	Search 'Add Image'
sl. item (1) (a) (i) name (bloc (ii) code	(2) :k letters)	Fiel	d Inve or Stati (J	stigato istical ISO) (3)	r (FI) / Officer	Fiel Sta	l Officer istical O	(FO) fficer	Senior (SSO)		Ŋ	🔁 Export PDF 🗸 🗸
(1) (a) (i) name (bloo (ii) code	(2) Sk letters)		(	(3)								Edit PDF
(a) (i) name (bloo (ii) code	ck letters)						(4	I) –		1		🚱 Create POF 🗸 🗸
(ii) code		+ ,								1		Comment
										1		🚯 Combine Files
(iii) signature				-		<u> </u>				1		扪 Organize Pages 🔺
(b) (i) name (blo	:k letters)									1		Delete, insert, extract and
(ii) code				Γ		1						rotate pages.
(iii) signature		<u> </u>		-	-	1					,	Try naw
2. date(s) of:		DD	Ν	(M	YY	DD	MN	1	YY	1		Compress PDF
(i) survey/in	spection									]		🔏 Redact
(ii) receipt										]		Protect
(iii) scrutiny										]		Adobe Sign
(iv) despatch												🔏 Fill & Sign
3. number of ad	ditional sheet(s) attached											🔓 Send for Comme
4. total time tak	en to canvass the schedule by the					1						
team of inves	tigators (FI/JSO)											Convert, edit and e-sign Pl forms & agreements
(in minutes) [	no decimal point]											Free 7-Day Isla
22	<ul> <li>(i) name (bloc)</li> <li>(ii) code</li> <li>(iii) signature</li> <li>(iii) signature</li> <li>(iii) signature</li> <li>(i) survey in</li> <li>(ii) receipt</li> <li>(iii) scrutniny</li> <li>(iv) despatch</li> <li>number of ada</li> <li>total time take</li> <li>team of invess</li> <li>(in minutes) [</li> <li>number of minutes)</li> </ul>	b)       (i) name (block letters)         (ii) code       (iii) signature         date(s) of.       (i) survey inspection         (ii) receipt       (iii) scrutniny         (iv) despatch       (iv) despatch         total time taken to canvasa the schedule by the team of unvestigators (FL/SO) (in minutes) [no decimal point]         innuber of advisors (FL/SO) in the team who	(i) (i) name (block letters)       (ii) code       (iii) signature       date(s) of.       (i) survey inspection       (i) receipt       (ii) scrittiny       (ii) scrittiny       (iv) despatch       number of additional sheet(s) attached       total ime taken to canvass the schedule by the team of mixes(gators (FL/SO) (in minutes) [no decimal point]       unumber of ny-settgators (FL/SO) in the team who	(i) (i) name (block lefters)           (ii) code           (iii) signature           date(s) of:           (ii) signature           (ii) signature           (iii) signature           (ii) signature           (iii) code           (iii) code           (iii) code           (iii) code           (iii) compt           (iii) contrany           (iv) despatch           total time taken to canvass the schedule by the team of avestigators (FL/SO) (in minuber) of codemal point]           innuber of movestigators (FL/SO) in the team who	(i) (i) name (block letters)           (ii) code         Image: Control of the second secon	b)         (i) name (block letters)           (ii) code	(i) (aname (block letters))	(i) (aname (block letters)         Image: Second Secon	(i) (aname (block letters)         Image: Constraint of the term who           (ii) code         Image: Constraint of the term who           (iii) code         Image: Constraint of the term who	(i) (name (block letters)       Image: Second	(i) (anate (block letters))       Image: Constraint of the con	(i) (anaue (block letters))       Image: Control of the

## (Refer Slide Time: 30:21)

3     numerity matcuto       4     total time laken to canvass the schedule by the team of investigators (FUSO) (in munities) [no decimal point]       5     number of unvestigators (FUSO) in the team who canvassed the schedule       6     whether any remark has been entered by [FUSO] (ii) elsewhere in the schedule       71/25C supervisivity officer (yes-1, no-2)       8				🖁 - 🖵 📮 🖉	-) (i) (14%, ·	🕒 2 / 16 🖡 🕙	(	Q	🗘 🕀 🖶
total time taken to carvas the schedule by the team of mvestigators (F1 JSO) in the team of mvestigators (F1 JSO) in the team who carvas due schedule     whether any remark has been entered by F1 JSO-spectra the schedule     for whether any revisity of the schedule     If JSO-spectra the schedule     If JSO-spec			^			anacieu	manori or agamonat sire	э.	
team of mvestigators (FI/SO)       imminutes) [no decimal point]         5.       mumber of mvestigators (FI/SO) in the team who canvassed the schedule         6.       whether any remark has been cancered by FI/SO supervisory         (i) in block 12/13         been cancered by FI/SO supervisory         (ii) elsewhere in the schedule         (iii) elsewhere in the schedule         (iiii) elsewhere         (iiii) elsewhere						e schedule by the	total time taken to canvase	4.	
(in minites) in o decunal point	Export PDF					0)	team of investigators (FI/J		
5       number of unvestigators (FL/SO) in the team who       Image: Comparison of the team who         6       whether any remark has been entered by the team who       Image: Comparison of the team who         7       (i) in block 12/13       Image: Comparison of team who is the team who is	Edit PDF		17			at	(m nnnutes) [no decinial p	<u> </u>	
6       whether any remark has been entered by F1 JSO supervisory officer (yes-1, no-2)       (i) in block 12/13       III         112       remarks by investigator (FUJSO)       III	Create PDF					SO) in the team who	number of investigators () canvassed the schedule	5.	
F[]JSO'supervisory     (ii) elsewhere in the officer (yes-1. no-2)     Schehule     Iii)       [12] remarks by investigator (FUJSO)     Iii)     Iiii)	Comment					i) in block 12/13	whether any remark has been entered by	6.	
[12] remarks by investigator (FUJSO)	Organize Pages					<li>ii) elsewhere in the chedule</li>	FI/JSO/supervisory officer (yes-1, no-2)		
[12] remarks by investigator (FUJSO)         Image: Constraint of the second s	insert, extract and pages.								
	Try now	,				ISO)	remarks by investigator (H	[12]	
	Compress PDF								
0	Redact								
	Protect								
	Adobe Sign								
4	Fill & Sign								
[13] comments by supervisory officer(s)	Send for Comme.					lcer(s)	comments by supervisory	[13]	
Cover In	err, eox aid e-sign i lorms & agreements								

Here is our questionnaire or schedule.

### (Refer Slide Time: 30:23)

![](_page_27_Figure_1.jpeg)

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)

![](_page_27_Figure_4.jpeg)

### (Refer Slide Time: 31:08)

![](_page_28_Figure_1.jpeg)

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)

Auto	Same 💷 🗟 🔌 🔧 🕫		6636	y75,25	1(1) · Ce	mpeti	illey Mod	۰.		٩	Search											Ramesh Jobs	untan (		1	o x
file	Home Insert Draw Page Layout	Formal	las Dat	a 1	Review	Ve	- 19	φ.																d 9	tare 🖓	Comments
	X Cut D Copy - B J U - H - Ø	. А' А'	H H H H		). 1 I I	终 Wie	p Text te & Cer	ter -	General		Conditional Farma	No No No	rmal_She	Normal	Bad Check (	લા	Good Explo	natary		set Delet	e Format	∑ AutaSun ∎fill +	- 27 Son 8	O Field	Arialy on	
	V Format Painter	Ξ.						÷.,			Fermatting + Table									· ·	÷	4/ Cer+	Filter	Select *	Data	
	coprote to the				Argune				s again					sters						Care			coursy		Nappa -	
189	* 1 × × µ					1.4					w.													. v	1 100	1.4
20	11 Percente in Mark 17-13	5	40	6		1	14	34			N.						Ų	P	ų	R	3		U	v	w	-
40	12 Remarks chewhere in Sch		60	- 1		2	77	22																		
41	11 Remarks chewhere in Sch		60	- í		1	78	78																	1	in the second
42	34 Black		- (4)			48	79 .	126																		
43	35 NSS					3.1	27	129																		
44	36 NSC					3.1	30	132																		
45	37 MULT					10 1	33	142																		
45 47	Sch. 25.0 : LEVEL - 02 (Block 3)																									
48										_																
47 11 50 D	L Item	Schedu Block	Item (	ice fol	Lengt	• •	te posi	108	Kemarks																	
51	1 Common-ID		in the second		_	34	1 -	34	Auto-dapăcated	_																
52	2 Level					2	35 -	36	'02' Generated																	
53	3 Filer					5	37 -	41	'00000' Generated																	
54	4 Household size	3	1			2	42 -	43																		
55	5 Whether HHD paid major share for childbirth expenses for any non-HHD	3	2			1	44 -	44																		
56	6 NIC-2008 five digit code	3	3			5	45 -	. 0																		
0	7 NCO-2004 three digit code	3	4			3																				
	1 Montabuld tone		4				<u> </u>	~																		
58	a structure type		1				53	53																		
59	9 Keapon	,	÷.			÷.	54 -	54																		
60	to socia group						55 -	55																		
61	11 Type of latrine usually used	,	8			4	56 -	57																		
62	12 Access to latence		9			5	SK	58																		
63	13 New many members use the latrine	3	10			1	59 -	60																		
64	14 Major source of drinking water	3	11			2	61 -	62																		
44	15 Arrangement of garbage disposal	3	12			i.	43	-																		
55	16 Primary source of money for cooking	- 1	11			2	4	40																		
	17 Was there sudden outbreak of	- í	14			î.	66 -	66																		
67	communicable disease in the community																									
68	18 Amount of medical insurance premium (Rs.)	3	15			8	67	24																		
69	19 Household usual commer expenditure (Rs.)						18	82																		
70	30 Black					44	83 -	126																		
71	21 NSS					3.1	27	129																		
-	layout75.25.0 (+)					-																				
Parts.																						8		<b>河</b>		+ 127
	<b>A A A</b>														144.5				-							22 -
	P Type here to search		0	R	۷.	*	6		estraction	4	NPILL Lechare23	- 44	Rantar Sel 15.1 -	C-10	erizante i	w.,	<b>1</b> 40	alay 11,796	09		26.0	^ 🧿 🕻	•	÷.	JAG 15-11	-2021 📆

You can just have a check on the layout file.

### (Refer Slide Time: 31:45)

AutoSave 🚥 🗟 🗇 - 🤆 - 😽 + 💷		datalay 75	258 (I) - Con	patibility Me	de •	P	Search							Ramesh	Interandam	•	-	0
ile Home Insert Draw Page Layout	form	las Data	Review	View 1	telp											d 9	are 🗆 C	Comme
X Gut Copy ~ S I 및 - ⊞ - S	- A' A'	× × 8	** # 5 5 8	Wap Text Merge & Ce	Gene riter - 188 -	nd . % 9 11 4	Conditional Forr Formatting * Tab	Normal_She	Normal Calculation	Bad Check (ell	Good Exploriat	- - 	Inset Delete	Format γ	uSum = 25 v Sot ur= Fite	7 O & Find &	Analyse Data	
Optoard IS Fert			Alignment		6	Nutber	5		Styles				Cells		Editing		Analysis	
• • 1 × - 4																		
A	с	DE	F	GH	1.	J	К	1.1	м	N	0	P Q	R	s	T U	v	W	1.8
36 NSC V METT				3 130	132													
Sch. 25.0 : LEVEL - 02 (Block 3)		_															Ro	por 44
srt. Item	Sched	ule reference	Leagth	Byte por	uition	Remarks												
80.	Block	Item Col.																
1 Common-ID			. )	1.1.	34 Au	to-daplicated												
1 Film				1 32 -	41 100	" Generated												
4 Household size	·	( i i		2 42 -	43	- clainte												
5 Whether HHD paid major share for childbirth expenses for any non-HHD	3	1		44 -	44													
6 NIC-2008 five digit code		3		\$ 45.	49													
7 NCO-2004 three digit code	1	4		50.	52													
8 Household type	3	( <u>s</u>		1 51 .	53													
9 Religion				1.54.	54													
10 Social group	3	7		1.55	55													
11 Type of latrine usually used	1	8		2 56 -	57													
12 Access to latine	3	9		- 58 -	58													
13 Hew many members use the latime		10		59 -	60													
14 Major source of drinking water	1	- 11		61 -	62													
15 Arrangement of garbage disposal	1	12		63 -	63													
16 Primary source of energy for cooking		13		2 64 -	65													
17 Was there sudden outbreak of	3	14		1 66 -	66													
18 Amount of medical insurance memium (Rs)																		
19 Household usual consumer expenditure (Rs		10		5 75 -	82													
30 Blank			4	4 83 -	126													
21 NSS				3 127	129													
22 NSC				3 130	132													
23 MULT			)	0 133	142													
Sch. 25.0 : LEVEL - 03 (Block 4)																		
1 Isyout75 25.0 (+)																		
after a state of the state of t															8 0	四		
			-		-		-	_	-							1.0		

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 givens 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)

![](_page_30_Figure_1.jpeg)

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)

![](_page_31_Picture_1.jpeg)

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)

🗐 Sata SE 15.1 - C. David States Alfred to	planary 2020 Palap Mehany NS 2014640 are	school and - 02 Bhock () dia					- 0 ×
File Edit Data Graphics Statistics	User Window Help						
😂 🖬 🏶 🗉 🖻 - 💷 - 🛃 - 🛃 5	3 🗇 O O						
Review T # X					X	Variables	1 # X
titer commands have 0	(8)					A Liter variables	here
# Command re		Conversion 1991-1912 deservices 11/				Nave	Label A
use "C\Uber\Studio #/Des		StateCorp				FSU	
		4905 Lakeway Drive				Segment	
	Special Edition	College Statice, Texas 77045 USA				\$55	
		800-STATA-PC http://www.stata.com				Household	
		979-696-4600 stata@stata.com				HouseholdSize	
		4.4-646-4601 ((88)				childishlip	
						NIC2008_6ve	
	Annual Annual Concentration					NC0204,thr	
						Household_t.	
						Religion	
						SocialGroup	
	Detecta is supported; see	ale unicode admice				Type of Jans	
	<ol> <li>Maximum number of variable.</li> </ol>	is set to 5000; see help set maximum.				Access_to_lat	
						How, many,	
	. use "C:\Users\Studio A\Desktop\Jan	ary 2022\Dr. Fratap Mohanty\MSS 75th Bealth\extr	otion/Level - 02 (Block 3).dta"			Anner Searce.	
						Primary store	
						Outherable	
						medical incu-	
						Properties	* *
				R		d Madahian	
						Nerse	Heusehold
						Label	
						3,91	and
						Format	5.01
						Value label	
						* Data	
						Liename	Level - D? (Block 7)-
						Notes	
							26
							7.92M
						Memory	54M
						Sorted by	1010
	Command						
Usen/Studio A/Desktoa/January 2022/Dr.	Pratao Mohanti/Stata15						OP NM OF
	0 m 0						1126
P Type here to search	0 🛱 🌘	🥙 🕐 📙 estados 🛛 🥬 NPIL, Lecture 👔	PowerFairt Std. Sata/SI 15.1	Cata_layout_Fit 1 detainy75.)	901- 🥥 2576 🔷 🖸 🛙	¢x ∯ 046	15-11-2021

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)

![](_page_32_Figure_2.jpeg)

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.

(Refer Slide Time: 36:19)

🚦 Statu/SE 15.1 - C/User/Studio N/Dealtop	gi January 2022/Dr. Pratagi Mohanty/MSS 79th HealthY acti	ctor/J.evel - 02 (Block 3).dta						- 0	×
File Edit Data Graphics Statistics	Uter Window Help								
🗟 🛯 🌚 🗍 🖼 - T - 🔍 🕄 🕄	3 0 0								
Review         T # ×           Interconversarials have         0           #  Conversarial         jrc           1         use "CEUben/Studie A/Des           5         bits/Studie A/Des	(8) // / (8) // / 15.1 Statistics/Data Analysis	Copyright 1985-2017 StateCorp LLC StateCorp					Variables Liter variables Name Househeld_t Ridicion	n here Label	^ <b>*</b> ×
2 BE PHILO FOU Segment 355	Special Edition	CollepsiteLon, Texas 77045 USA 800-STATA-PC http://www.stata. 978-686-4600 stateBstata.com 978-686-4601 (fax)	con				Kergen SocialGroup Type_of_lati Access_to_lat How_many		÷
	Anna agus Annas Anna agus Annas Annas - A banas						Major, source Anangement Primary, sour OutbreakCo medical_insu		
	Notes: 1. Unicode is supported; see 2 2. Maximum number of variables nee TC:\NearalStrutio AlDasktonhJaco	elp unicode_advice. Is set to 5000; see help set_maxwar.	aalthlastractics/laval - 50	(North 1) dias			Househeld_U NSS NSC MULT		
	br HHID PST Segment SSS Honsehold	ary proposition fractage measury was not a	eestooleetteolloolloest - 15	DELOCK 31.00%.		4	HHD weight,s weight,sc		_
							Properties	HHD HHD Sile	8 ×
							Costa     Edename     Label     None     Variables     Observations     Saic     Memory     Sorted by	26 11),821 7430 840 100	nck (), d
	Command Levid MRID								
CitheriStudio #DextrayUsrway 2022.0.	Putap Mohamy/Stata15	🛷 💽 📙 estadion 🏼 🤌 MRI,	lacture. 👔 PowerPoint Std.	Statu/SI 15.3 👌 Data_kryou	(FL 👖 denday 75,250 L.	🥥 asc. 🔺 🧿 🛙	¢ d <sup>4</sup> 040	CAP NUX 6 1128 6 15-11-2021	A OR

So, for us id variable is well defined, but we need to know that whether that id is actually uniquely identified or not. That is one of the important aspect we need to understand either you check for id to be uniquely identified you have to set the command called isid. So, isid then id variable first HHID will check first with HHID just HHID since we have defined and then enter.

(Refer Slide Time: 36:54)

💐 Statu/SE 15.1 - C/User/Studio N/Dealtop	(January 2022)Di-Pratap Mohanty(NSS 75th Health) extra	tan/Levi - 02 (Block I) dia		-	0 X
The tall talls traphet statistics	the mater hep				
	1000				
Image         Image         Image           (Intermentation         (Intermentation)         (Intermentation)           (Intermentation)         (Intermentation)         (Intermentation)           (Intermentation)         (Intermentation)         (Intermentation)           (Intermentation)         (Intermentation)         (Intermediation)           (Intermentation)         (Intermediation)         (Intermediation)           (Intermediation)         (Intermediation)         (Intermediation)	(0)     (1)     (	Copyright (181-2017 fixedory 102 Santalan Santal	b.	iddes Etter vander her New Lat FSU SSS Household Son Antibiotics No2200 (Ar., No2200 (Ar.,	743
	. use "C:\Deers\Studio A\Desktop\Janu . br HHID PSJ Segment SSS Honsehold	ry 2022/Dr. Freisp Hohanty/MSS 25th Bealth/astraction/Lawel - 62 (Block 3).dta*		Anangement Anangement Primaty.pour OutbreakCo	
	. Leid WHID			medical, insu.	
			Pro	perties	
				a a	
				Jari Japi Iornat I Iornat I Iornat I Iornat I Ioraane I Ioraalis I Ioraalis I Ioraane I Ioraalis I Ioraane I Ioraalis Ioraalis I Ioraalis Ioraalis Ioraalis I Ioraalis I Ioraalis I Ioraali	al) 584 audi - 67 (Block I) 26 113,621 500 500 600
Ciliberi Gudia // Deltopi lanuary 2022/br. 10	Connand Land 1919 Segment 555 Bouwehold huog Mehaniy/Surat5 O 12 (9)	⑦ ● ■ marker ● 1991, Lenver、 ● Powerfurt 36. 第300(971) - ● Sec. Jour, 10. 1月 000(97,75).	v ∎ ⊕ asc ∧ ∳ ₽ ≇ •	e de lous ,	CAP NUM D 1128 5-11-321

Now, if there is no error or red line displayed; that means, this variable is perfectly uniquely identified in this particular data set for the observation. Otherwise, you can also check the

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)

Q Star 20 31 - Classer (Studies Richardsparkars) 2023/b. Analy Mitheney Mich 200 International Levil - Q (Stock 2 data Star 2 data Charter Starters Start Starters)		- 0 >	×
Image: International State     Image: International State     Image: Ima	<ul> <li>Veridaes</li> <li>The condition</li> <li>Name</li> <li>Name</li> <li>Religion</li> <li>SocialGraps</li> <li>Type of Jahr.</li> <li>Access (N, M.L.</li> <li>Access (N, M.L.</li> <li>Arrangement.</li> <li>Arrangement.</li> <li>Arrangement.</li> <li>Arrangement.</li> <li>Arrangement.</li> <li>Arrangement.</li> <li>Arrangement.</li> </ul>	T 4 horr	
<ul> <li>Minimi         <ul> <li>Distribution in experiments are high selected Affine.</li> <li>Restame states of verifables (s set to 100); and high selected selected in the set (based).</li> <li>were "C (Minimizitation Affinishinghomeony MERSIN: Finding MediantyAffit 700 Reside Selected and the set of the selected selected selected selected selected selected selected selected and the selected select</li></ul></li></ul>	rredukt,mou. Household,U NSS NSC NULT HHID weight,ss weight,ss		D
- Intel WP Support SS Research	Properties Properties Provide Provid	9 1040 5/8 11/00 25 1111/21 2/804 1010/1 2/804 1040	1.8
Connucle arts MBD Claum-Sudar Ediningsiumy 2026. Angel Manay Naval D Spectres to Sach	9 2 de de 100	CAP NUM 11-28 15-11-2021	01

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)

![](_page_34_Picture_5.jpeg)

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

(Refer Slide Time: 37:47)

M & A & . T . Q . G	3 0.0			
iev Tax	<i>a</i> 1		<ul> <li>Variables</li> </ul>	+ a :
Litter commands have			1 litter vanables P	
Command _rc		Copyright 1905-2017 StateCorp LLC	Nave	Label
use "C106en/Studio A/Des	Statistics/Data Analysis	StateCorp	Household_t	
br HHID FSU Segment SSS	August Barris	4905 Lakeway Drive	Religion	
sid HHD	special Edition	College Station, remaining to a	SocialGroup	
rid FSU Segment SSS Meus		979-686-4600 stataBstata.com	Type_of_lats	
son mile	1	979-696-4601 (fax)	ACCEST-00_ML	
save, reprace			How many	
	It was not an angle of the second		happy jugarde	
	the set of the second s		Primary store	
	and the second s		Orthonal (	
			medical inter-	
	Notes:		Household U	
	<ol> <li>Unicode is supported; see h</li> </ol>	elp unicode_advice.	N53	
	<ol><li>Maximum number of variables</li></ol>	is set to 5000; see help set_maximat.	NSC	
			MULT	
	. use "C:\Users\Studio A\Desktop\Janu	ary 2022\Dr. Fratap Mohanty\WSS 75th Health\extraction\Level - 02 (Block 3).dta"	HHD	
	br HID FSI Sement SSS Banashold		weight, ss	
			weight_sc	
	. 1#14 HEID			
			Properties	
	. 1818 PSU Segment SSS Ecusebold		8 F.F.	
	. sort HHID		# Variables	
			Name	HHD
	. save, replace		Label	
	file C:\Users\Studio A\Desktop\Januar	y 2022/Dr. Pratap Mohanty/NSS 75th Health/extraction/Level - 02 (Block 3).dta saved	3,04	18-9
			Formal	5.91
			Value labor	
			4 Date	
			Label	
			Nater	
			Variables	26
			Observations	113,623
			Sat	7.93M
			Memory	54M
			Sorted by	HHD
	former d			
	Command		1	
	CIEAT	I		

This may be our master file, we need to check we have saved and replaced it. Now we are going to open the using file and we need to check their unique idea as well.

(Refer Slide Time: 37:58)

(8)	Copyrapht 1985-3017 Standarg 132 Standarg Will Lawry From Gilling Form THE USL Colling Form T	<ul> <li>Variables</li> <li>Stars consider here</li> <li>Nere an enders to photoes to photoes</li> </ul>
(8) 	Copyright 1985-2017 Streading 122 Denoting Scillary Rev File College Revents, Rever 1984 1934 1997 1997 1997 1997 1997 1997 1997	▲ Ether canada here Name Label Deck are no items to alw
	Capital (19-201) Sundary LLC Sendary 490 Laney Part College Senter, news 1794 SBA 500-35524 March 1997 (2004) SBA	Name Label There are no items to also
Statistics/Data Analysis Apecial Edition	Seascing 485 Lakery Drive College Station, Texas 77645 003 00-9702147-0 Nata Nata College Station College 10-9702147-00 Nata College Station College Station College 10-9702147-00 Nata College Station College Station College Station College 10-9702147-00 Nata College Station Co	There are no items to alk
Special Edition	4405 Lakeway Drive College Station, Teama 7745 USA 000-17178-00 http://www.state.com	
Special Edition	College Station, Fexas 77945 UDA 800-STATA-PC http://www.stata.com 954-664-460 statistata.com	
	979-696-4600 stataBatata.com	
	979-696-4601 (fax)	
<ol> <li>Unicode is supported; see &gt;</li> </ol>	elp unicode advice.	
2. Maximum number of variables	is set to 5000; see help set maximar.	
see "C:\Deers\Studio A\Deskton\Jaou	ary 2022\Dr. Fratan Mohanty\NES 75th Bealth\estraction\Level - 02 (Block 3).dta*	
a mith Poll Comment SSS Researcheld		
and meth		Properties
sid PSU Segment SSS Household		a
eort MHID		< Variables
		Nama
save, replace	- MANINE Reason Mohaman MIT Mah Destah) established - Al (Black 1) des sound	Later
e crossistorato attestopositat	y avaaluit fiscap monentyinda toon messentestatutteves - va Intoon aytusta baven	Lornat
llear		Value label
		Notes
		* Outa
		Topcane Label
		Notes
		Variables 0
		Observations 0
	<ul> <li>March &amp; Marchael Marchael</li></ul>	<pre>Number 4 Particular Number 4 Number Number 4 Number 4 Numb</pre>

So, let us open another file we have now cleared the data saved in Stata.

(Refer Slide Time: 38:05)

![](_page_36_Figure_1.jpeg)

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

(Refer Slide Time: 38:10)

			A COMPANY AND A COMPANY	_
· · · · ·	(A)		A VALUES	
formed and a				i tana
commana jrc		Copyright 1905-2017 StateCorp LLC	interes and a second	Line
to HHD FSJ Second SSS	Statistics/sats Analysis	4405 Lakeway Drive	broitainel	
sid HHD	Special Edition	College Station, Texas 77045 USA	no. of times.	
isid FSU Segment SSS Hous		S00-STATA-PC http://www.stata.com	Whather pre-	
soft HHID		979-696-4600 stats@stats.com	child_bith_e.	
save, replace		3/3-636-4601 (E88)	communicab	ê.,
hai			soffered_chr.	
use "Cillben/Studio AlDes	term agent formation		suffering_any	1-
	transmitter in the second		alment_day_	
			covered by a	h.,
	Notes:		Reporting_or	
	1. Unicode is supported; see	lp unicode advice.	103	
	<ol><li>Hesimum number of verieble</li></ol>	is set to 5000; see help set maxwar.	TIM	
			HHD	
	. use "C: (Dsers(Studio A(Desktop)/260	ry 2022/DF. Fratap Monanty/MSS 2505 Bealth/extraction/Level - 02 (Block 3).dta-	P. P.O	
	br HIID FSI Segment SSS Household		Wineight, 15	
			weight_sc	
	. isid HHID			
	and BE formers off Broucheld		Properties	
	· · · · · · · · · · · · · · · · · · ·		8 · · · ·	
	. sort HHID		* Variables	
	. save, replace	ANTING Reason Makanani MAR Mak Realshi and and initianal . All (Black 1) day sound	Los	
	THE COMPLETIONS ADDRESS OF COMPLETE	source: Freedy Ministry and Instrumentation Const of Direct Street Bases	Lorent	
	. clear		Value label	
			Noter	
	. use "C:\Users\Studio A\Desktop\Jen	ry 2022\Dr. Fratap Mohanty\WSS 75th Bealth\extraction\Level - 03 ( Block 4).dta"	* Data	
			Libel	
				21
			Observations	
				35.48M
			Memory	9644
			Sorted by	PD
	Command		,	
	191d FID			

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)

. 10			Veriables	1.0
(decomposite how )	(8)		A 100-000	
			Total Cardina	
Command Jrc		Copyright 1905-2017 StateCorp LLC	Nava	Label
La	Statistics/Data Analysis	platacorp	UNA ANDA	
ER PPHO PSO SEGMENESSS	Special Edition	College Station, Texas 77045 USA	nonpranten	
isid FSU Seament SSS Hous		E00-STATA-PC http://www.stata.com	Wether ere	
sort HHD		979-696-4600 stata@stats.com	child both e	
save, replace		979-696-4601 (fax)	communicab	
dear			suffered_chr	
use "C-Usien/Studio A/Des	from all the second		suffering_any	
aid PD			ailment_day	
			covered_by_a	
			Reporting_of	
	Notes:	No unicode educe	N55	
	2. Maximum number of variable	is set to 5000; see help set maxwar.	NSC.	
			MULT	
	. use "C:\Dsers\Studio A\Desktop\Jan	ry 2022\Dr. Fratap Mohanty\#55 75th Bealth\extraction\Level - 02 (Block 3) dta"	HHU	
			PIO	
	. br HHID PS7 Segment SSS Bonsehold		weight u	
	Laid HITD			
			Constant of Consta	
	. 1sid FSU Segment SSS Ecusehold		risperies	
			4 Mariablas	
	. sort HHID			
	any replace		Label	
	file C:\Users\Studio A\Desktop\Janua	2022\Dr. Pratap Mohanty\N55 75th Health\extraction\Level - 02 (Block 3).dta saved	501	10/11
			Format	
	. clear		Value label	
	nee 10 (Means) Studio (Mashton) Jac	en 1999)Br. Bratan Mohantu/MES 19th Basithiantrantion/ienal - 61 / Black Al dta"	Fisher Data	
	. use .c. (users(studio A(usskiop(ren	th horsens and workershows over meeter/estimation/manet - ro ( ploce e) are.	- Data	
	. teid PID		Label	
			Fister	
			Variables	28
			Observations	
			Sas	35.48M
			Memory	964
	1			
	Command			
	ciea			
		I		

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)

![](_page_38_Figure_4.jpeg)

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)

![](_page_39_Figure_1.jpeg)

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

(Refer Slide Time: 42:20)

Normality         0				× W	wiables	1/
amount of the sector of the	I die commande have	(1)			Litter canables b	
Terminant Aller and Statistic St	Comment				Neme 1	abal
<pre>minimum manufacture manuf</pre>	contrates (1)		Copyright 1905-2017 StataCorp LLC		Muncheld a	100
<pre>infit definition of the second s</pre>	to HHID FSJ Seisnerit SSS	statistics/cata Analysis	HateLorp 4905 Lakeway Drive		Relains	
Million	ind MHD	Special Edition	College Station, Texas 17045 USA		Sociations	
and B index and B	isid FSU Segment SSS Hous		200-STATA-PC http://www.stats.com		Type of Jati.	
an angin	soft HHD		379-696-4600 stata@stata.com		Access to lat	
Max     Max     Max       Max     Max <td< td=""><td>save, replace</td><td></td><td>179-696-4601 (fax)</td><td></td><td>How, many</td><td></td></td<>	save, replace		179-696-4601 (fax)		How, many	
ar ('bit hadked). We want the set of the se	clear				Mejor, source	
<ul> <li>High and the set of the</li></ul>	use "C\Usen\Studio A/Des	from agent formation			Anangement	
are "Considered and a set of a	aid PID	summer or the boundary			Primary_sour	
ar Classified Add Add Series Constraints and a series of the series of t	dear				OutbreakCo	
<pre>next</pre>	use "CI/Usen/Studio A/Des	J			medical insu	
<ul> <li>Anome under of weighten is bein to DDD, we high on games.</li> <li>Anome under of weighten is bein to DDD, we high on games.</li> <li>We then the DDD is th</li></ul>		Notes:	le unitede edute		Household_U	
<ul> <li>as - C (Marcy/Mark &amp; Markky/Lossey H17): Parky Marcy/MS 710. Bull/Market/antial(an) - 12 (Block 1) .84*</li> <li>be HE 705 Segars 33 Rounds18</li> <li>be HE 705 Segars 33 Rounds18</li> <li>as - W (Block 1) .84*</li> <li>as - W (Block 1) .8</li></ul>		<ol> <li>Maximum number of variable.</li> </ol>	is set to 5000; see help set maxwar.		N53	
- ar C (hard) (hat & Kakata) (hard) (1975) - Freing Manary (1975) This Mail (hard) (1976) - 10 (files 3) - 60° - W HD 76 Sequence 18 Annahold					NO.	
<ul> <li>bit 202 976 Seguest 155 Securit-15</li> <li>bit 202 976 Seguest 155 Securit-15</li> <li>bit 202 976 Seguest 155 Securit-155 Secur</li></ul>		. use "C:\Users\Studio A\Desktop\Jan	ry 2022\Dr. Pratap Mohanty\WSS 75th Bealth\extraction\Level - 02 (Block 3).	fta"	HULT	
bi HE DF Market Ma Mandel Mandel (1997) int of HE Segment MA Ma					which is	
Lind HD		. Br Hild F53 Segment 535 Bonzehold			weight or	
intel 700 Seguent SAS Researching       intel 700 Seguent SAS Researching     Perform       intel 700 Seguent SAS Researching RESPECt Proofs Researching SAS TAS Residuation Security SAS TAS Residuation Security SAS Res     Perform       intel 700 Seguent SAS Researching RESPECt Proofs Researching SAS TAS Res     Perform       intel 700 Seguent SAS Res     Perform		. Leid HEID				
- aret ETB		. 1414 FSU Segment SSS Equebold		P	opertes	
ent HB						
nere, rejisi 1:0: Collected Manage January 2007/06, Foreig Manage 2007 31% Machine Institution 2:0 (Bank B. J. An over 1 and 2 and		. sort HHID		·	Variables	
ner, gringen Strateg Manary 2023 fer, Frange Manary 2023 Talk Backbarrenten Lateffer 1 - 23 (Back B), dia aved in a set (Second Manary 2023 Second					Tabel	
- deer		. save, replace	MATLER Barrow McConstitute That Barlabi assessed on Faunt - Al (Black 1) de	i unut	Los	
- disar - star - C. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - Periop Relativities 1703. Build/Anteresticalizes) - 10 ( Nonk 6) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - 20. Mana (Shank 7) disar - star - 20. Mana (Shank Albackup/Joseny 1973): - 20. Mana (Shank 7) disar - star - 20. Mana (Shank 7) disar -		The conversion atteactprovide	avaruat. Franky Monancy and Tana Means Version Level - of Tanan 17.00	1 saves	format	
<ul> <li>ser -C. (here/thetis (bestep/beaury HITM: Prolog MassiyKK TM: Build)setration(lens) - 10 / Hosk II / Ha<sup>2</sup></li> <li>test FD</li> <li>diser</li> <li>as -C. (here/thetis (bestep/beaury HITM: Prolog MassiyKK TM: Build)setration(lens) - 10 / Hosk II / Ha<sup>2</sup></li> <li>diser</li> <li>as -C. (here/thetis (bestep/beaury HITM: Prolog MassiyKK TM: Build)setration(lens) - 10 / Hosk II / Ha<sup>2</sup></li> <li>diser</li> <li>as -C. (here/thetis (bestep/beaury HITM: Prolog MassiyKK TM: Build)setration(lens) - 10 / Hosk II / Ha<sup>2</sup></li> <li>diser</li> <li>as -C. (here/thetis (bestep/beaury HITM: Prolog MassiyKK TM: Build)setration(lens) - 10 / Hosk II / Ha<sup>2</sup></li> <li>diser</li> <li>as -C. (here/thetis (bestep/beaury HITM: Prolog MassiyKK TM: Build)setration(lens) - 10 / Hosk II / Ha<sup>2</sup></li> <li>diser</li> <li>diser</li> <li>diser</li> <li>diser</li> <li>diser</li> <li>diser</li> <li>diser</li> </ul>		. clear			Value label	
- an "C Marking/Salary 2017/36. Finding Maanty 2017/36. This Sec 2018 And 2					Notes	
1 Ent F10 1 E		. use "C:\Dsers\Studio A\Desktop\Jan	ry 2022\Dr. Fratap Mohanty\WSS 75th Bealth\extraction\Level - 03 ( Block 4)	-414" ·	Data	
Inform					Liename	Level - D? (Blo
- Alase - Alase - an *C (hand)(table A)(headlap)(headry 1017)(hr. Parlog Weadly(1015 710), Berl 10(heat)(ant)(ant) - 10 (Block 1), Ala* 		. 1818 P10			Label	
see "C (Search) Sales Albertop (Josevy 303) Siz Tracey Mekery 303 Sizh Sezi Salestretten (Josev) - 52 (Search) Sizh Sizh Sizh Sezi Salestretten (Josev) - 52 (Search) Sizh Sizh Sizh Sizh Sizh Salestretten (Josev) - 52 (Search) Sizh Sizh Sizh Sizh Sizh Sizh Sizh Sizh		. clear				
- un "C Unarry (State Alberty State Alberty State Freide Meanty State State Analysis of State Alberty State St					Observations	
- Many Milling		. use "C:\Users\Studio A\Desktop\Jan	ry 2022\Dr. Pratap Mohanty\MSS 75th Bealth\extraction\Level - 02 (Block 3).	fta"		7.92M
· Book ref					Memory	64M
Amond I					Sorted by	1010
Connard						
Command						
. official and a second s		Provenue				
		command			N	

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)

![](_page_40_Figure_3.jpeg)

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)

العام المراجع           معاد المراجع           معاد المراجع	an a
Normality         Normality	atel
Note         1           National Activity         National Activity           National Activity         Nationactivity           National	abei
1000 - 451/10         Normality           44 (1) 2010 - 451 (100 - 451 (100 - 12) (100 - 1) (100 -	
لا بالله الله الله الله الله الله ال	
در المراقع ال المراقع المراقع ال المراقع المراقع الم المراقع المراقع المراق المراقع المراقع المراقع المراقع المراقع المراقع الم	
art th Sills are fold any Annuel.     The Sills are fold any Annuel.     The Sills are fold any Annuel.     Sills are	
JUDYSF. Proce Waaryykk 200. Buddateretinkiemi - 12 (Biok. 3). An Wildley USZNY, Proce Waaryykk 200. Buddateretinkiemi - 12 (Biok. 3). An Wildley USZNY, Proce Waaryykk 200. Buddateretinkiemi - 12 (Biok. 1). An and Navyyk, Proce Waaryykk 200. Buddateretinkiemi - 12 (Biok. 1). An Wildley USZNY, Proce Waaryykk 200. Buddateretinkiemi - 12 (Biok. 1). An Wildley Wildley USZNY, Proce Waaryykk 200. Buddateretinkiemi - 12 (Biok. 1). An	
JBDDE Peop Many2013 THE RecitAterestication - 12 Allos 3: AL* USERS. Note: The sequence of the second sequence of	
HURL: Foreg Roberg1055 1105 Residuencestablem] - 10 (Rise 1)-far averaf	
Handbox, New York, State	
HERDY, Tesing Minington 101 Miningtonesesial.lengl 2 b) ( Name 17, 45°         Samity 2           HERDY, Tesing Minington 101 Miningtonesesial.lengl 2 b) ( Name 17, 45°         Samity 2           HERDY, Tesing Minington 101 Miningtonesesial.lengl 2 b) ( Name 17, 45°         Samity 2	
Surface Second S	
Starting         Starting	
Annu Su. NULLY, Parep Roberty 2013 Tell Reality Association (Jacob - 4) dilare 1), dia annue Paragona. Paragona. Paragona. 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue Paragona. 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue 2017 Dr. Parep Roberty 2013 Tell Reality Association (Jacob - 4) (Real 1), dia annue 2017 Dr. Parep Roberty 2013 Tell Reality 2014 Dr. Parep Roberty	
121-Dr. Proop Riberty/052 Tel: Buildhersetastialized - 12 (Biol: 2).db aread 222-Dr. Proop Riberty/052 Tel: Buildhersetastialized - 12 (Biol: 2).db aread 222-Dr. Proop Riberty/052 Tel: Buildhersetastialized - 12 (Biol: 2).db 222-Dr. Proop Riberty/052 Tel: Buildhersetastialized - 12 (Biol: 2).db 2222-Dr. Proop Rib	
UDDr. Freep MontpUDD 10h Registration(Leve) - Di Ubjer 1. As avend         Magazani.           Name, yes         - Di Ubjer 1. As avend         Ramputo.           Name, yes         - Di Ubjer 1. As avend         Name, yes           NUMDr. Freep Manapulti 10h Registration(Leve) - Di Ubjer 1. As avend         Name, yes         Name, yes           NUMDr. Freep Manapulti 10h Registration(Leve) - Di Ubjer 1. As         - Di Ubjer 1.         Name, yes           NUMDr. Freep Manapulti 10h Registration(Leve) - Di Ubjer 1.         - Di Ubjer 1.         Name, yes	
Status         Print Managing Tab. Buildharman Status         - 10 (Status 1). An anna A         Print Managing A           Status         - 10 (Status 1). Anna A         - 10 (Status 1). Anna A         - 10 (Status 1). Anna A           Status         - 10 (Status 1). Anna A         - 10 (Status 1). Anna A         - 10 (Status 1). Anna A           Status         - 10 (Status 1). Anna A         - 10 (Status 1). Anna A         - 10 (Status 1). Anna A	
Uniter Friedy Roberty Die Seiten Artificierie - 10 ( Robert 1), 453 4993 - Ordered C Constanti, - Constant	
. District, Perkap MenarySHS 20th Resitablestration(Level - E) ( Hork 4)-Ra* Parate Barter - E	
. HITUP: Foody Holanty Mill Tith Hallblottration(Lovel - 13 ( Mink (), 664" Population - Populat	
2022/Dr. Fratap Mohasty/USS 75th Bealth/astraction/Level - 83 ( Block 4).46a* Poppetes	
8 · · ·	
<ul> <li>Variables</li> </ul>	
Name	<2 selected+
Label	
2022/DF. Fratap Mohanty/MSS 75th Health/extraction/Level - D2 (Block 3).dts"	
Format	
AlDeskteplJannary 2022[Dr. Pratap Mohanty(NSS 75th Bealth)estration[Leve] - 02 (Block 3).dta*	
Notes	
* Data	
D Constant	
3.023 (merge=3) Note	
Looking	26
Observations	
Sat	7.92M
Memory	6M
2022/00. Pratap Mohanty/MSS 75th Health/extraction/Level - 67 (Block[2].dta" Solid by	HHD
A) (7) (3)	Bestophanny 1973/F. Prog Malary/MI 110. Ba(10)astraction(Javel - 0) (Block 1), Au <sup>2</sup> المنطق           db.         المنطق         المنطق           db.         (Block 1), Au <sup>2</sup> Non           Higher 1, Au         (Block 1), Au <sup>2</sup> Non

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)

N 🕾 (] 🛪 : T : 🕄 : 🕄	3 0.0		
ev Tax	<ol> <li>Maximum number of variables is set to 5000; see help set_maxvar.</li> </ol>	<ul> <li>Verables</li> </ul>	
l iter commands have	. use "C:\Deers\Stadio A\Deektap\January 2022\Dr. Fratap Mohanty\MSS 75th Health\extraction\Level - 02 (Block 3).dts"	filter variables 2	
Command _rc		Nava	Label
use "Cilibies/Studio A/Des	. br HHID PST Segment SSS Household	Relation_to_h	
bi HHID FSU Segment SSS	. Leid HHID	Gender	
inid FSU Seament SSS Hous		Marital status	
son HHD	. isid FSD Segment SSS Bonsebold	General_educ	
save, replace	4447 3010	UNA_status	
clear		hospitalized_L.	
use "Cillbien/Studio AlDes	. save, replace	no_ol_times	
oid PD	file C:/Umers/Studio A/Desktop/January 2022/Dr. Pratep Mohanty/MSS 75th Realth/extraction/Level - 02 (Block 3).dts saved	Whether pre_	
use "C\Usen\Studio A/Des	, clear	chiej 600 (E.	
merge 1 m HHID using "C		suffered chr	
dear	. use "C:\Deers\Studio A\Desktop\January 2022\Dr. Fratap Mohanty\MSS 75th Bealth\extraction\lavel - 03 ( Block 4).dta"	suffering any	
use "Cillbien/Studio A/Des	1414 375	aiment_day	
marge 1:m HHD using 10		covered_by_a	
	. clear	Reporting_of	
	nas "C'illeara)Studio AlfasktoniJannary 20030r. Tratas Mohanty/MES 25th Baalthlaytraction/Javal - 00 (Block 3) dta"	5 mm	
	. One c. Constational a Constation of Association and Association and Association (Association of Association of As	D.	
	. merge 1:m HEED naing "C:\Users\Studio A\Deaktop\Jannary 2022\Dr. Pratap Mohanty\HESS 75th Bealth\estraction\Level - 02 (Block 3).dta"	Properties	
	Result # of obs.	B + -	
		* Variables	
	not matched 0	Label	, see all
		Tor.	byte
		Lormat	9,710
	- clear	Value label	,marga
	use "Cilleers)Stodio Albeskton)Jaonary 2022107. Fratan Mohanty1855 71th BealthhestractionLawel - 02 (Block 3).dts"	1 Data	
		Liename	Level - 02 (8)
	. merge 1:m HEID using "C:\Deers\Studio A\Deaktep\January 2022\Dr. Pratap Mohanty\NSS 75th Bealth\estraction\Level - 03 ( Block 4).dta"	Label	
	tents distant	Noter	
		Variables	45
	not matched 0	Sat	51.73M
	matched 555,352 (_merge==3)	Memory	9614
		Sorted by	
	Commissé		
	London Contra Cont	,	
	re "esche		

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 other 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.

143	. Isid #SD Segment 655 Rosebold	<ul> <li>Veriables</li> </ul>	
er commands hava		🔹 🔥 Eiter variables here	
ommand _rc	, set min	Name Lab	el
e "Cillhen/Studio A/Des	, ave, replace	Relation to h.,	
HHID FSU Segment SSS	file C:\Umere\Studio A\Deektop\January 2022\Dr. Pratep Hohanty\NS5 75th Health\extraction\Level - 02 (Block 3).dts saved	Gender	
1 HH0		Age	
FSU Segment SSS Hous	- clear	Merital_status	
HHD		General_educ	
e, replace	. Des C. (Destalations of Americal AntiAnt. Lands Boowerk Area vie Bestrukerischen/mant m ( Brook 4) - and	URA_status	
	. Isid PID	hospitalized [	
C:Uses/Studio A/Des		no_of_times	
10	- diear	Whether pre	
		child_both_c	
C10ses/Studio A/Des	. see "C:\Userelstonio A\Desktop\cansery FF27\FF. FFstap RonandyUmss F2th meminivextraction\Lavel - 62 (Block 3).dta"	communicab	
pe I in HHID using "C	. merme in HEID saint "Civiners/Studio Albeston/Jappary 2029/Dr. Fratan MohantwikiS 75th Bealth/setration/Jappal - 02 (Block 3).dts"	suffered_chr	
· · · · · · · · · · · · · · · · · · ·		whening_any	
Croser states water.	Result # of obs.	ament_day	
ge on new using		Emotion of	
ioyi .	not matched 0	PD	
	satoled 113,823 (merge=1)	A (1973)	
		C.	
	. clear	Annuality	
		ropertes	
	. use "C:\Deers\Studio A\Deektop\January 2022\Dr. Fratep Mohanty\WSS 75th Bealth\extraction\Level - 02 (Block 3).dta"	d Muddhine	
		Nerve	
	<ul> <li>medde 1:B WHID SKIDE -C.(Neeks/stD2D0 V/neeks/stD2D1 North/ http://www.skithectico/neeks - 0.1 (Block 4) gra.</li> </ul>	Label	
	Result # of obs.	tor b	151
		Format 5	210g
	not matched 0	Value label	manga
	matched 555,352 (_merge==3)	Noter	
		* Data	
		Liename	
	· ra Tantka	Label	
	merge Freq, Percent Cum,	Variables of	
		Observations y	
	matched (3) 535,352 100.00 100.00	Sat 5	7.73M
		Memory 9	644
	10141 535,352 100.00	Sorted by	
	Command		

(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.