Handling Large-Scale Unit Level Data Using STATA Professor. Pratap C. Mohanty Department of Humanities and Social Sciences Indian Institute of Technology, Roorkee Lecture No. 21 Extraction in Stata-I

Friends once again I welcome you to my lecture on Handling Large-Scale Data Using Stata. This is as part of NPTEL MOOC module is sponsored by MHRD for our better use of data and better use of pedagogy in different set ups. Here we tried our best to enable you how to handle the unit level data with Stata and that too how to also interpret the data using Stata. This particular week is targeted for understanding some important intricacies of Stata handling of the unit level data. So, we have accordingly titled as handholding of unit level data.

(Refer slide Time: 01:18)



So, let me proceed for understanding the extraction as part of handholding. In these two lectures we are going to discuss the detailed information of data extraction and we will try our best today to stick to NSS or even economic census database and other data you can also explore. We will guide you in between for other databases wherever it is there.

So, let me just introduce you what do you mean by the extraction. Stata has its own format for storing datasets. We need to understand this background very clearly before extraction. Stata saves it in .dta files as I already guided you. These are highly convenient for Stata users, because

Stata can use them immediately without any need for interpretation. But many a times large scale unit level datasets are not available in Stata format. They are available in different formats or in, simply in raw formats. Raw data can be defined as generally an unstructured or unprocessed data.

(Refer Slide Time: 02:36)



So, data extraction is needed to retrieve the unstructured data for further data processing and analysis. Usually large scale datasets are available in excel format or Excel file that is in .xls or xlsx and also in ASCII text data format. If your data is originally in Excel or other format, you need to prepare the data before reading it directly into Stata. So, you need to prepare it as per Stata requirement.

(Refer Slide Time: 03:05)



So, what do you mean by ASCII data, it is called American Standard Code for Information Interchange. It is a set of codes of 256 characters used in computers. It includes alphabets that is A-Z in capital or in small, numbers from 0 to 9, control characters like enter and shift these are also there and also graphic characters. The ASCII text has raw data and/or delimiter. We are going to guide you on this in our next slide.

There are in general 3 groups of ASCII text depending on the delimiters which delimit the data or separates the data from one byte to another one. We are going to guide it clearly. So, these 3 are the following, free format, delimited format or fixed format. In this particular lecture we are going to guide you the handholding aspects of free format and delimited format. In the next lecture we are going to guide you the fixed format with the help of the core raw or the ASCII data available in India. What do you mean by free format data? Please do not miss any single point from our PPT. This is going to guide you very clearly.

(Refer Slide Time: 04:53)



Free format ASCII text separates data items using a space. Note it very clearly that space is important in this case. Like in the example dataset which we have put it here for your understanding, the entry of data is delimited by space, isn't it? Here the space is different. There is different from one entry to another entry. So, this format is simple and intuitive enough to be used for small dataset.

So, if there is a small data, then generally they go for this, otherwise the other formats are important. If your data are in free format, which I just guided you, with variables separated by blanks or comma or tabs, you can use the command called infile, infile command will guide you to extract the data to your Stata format. Like if it is in free format and which you can understand by the space separation between the variables.

The command here is infile, this is very important, then you wanted to extract the variable, that there are so many variables out of that for my use, my research paper, I want variable 1 to 5 or maybe 1 to 10 whichever are important, you have to take the name of that particular variable, this variable, that variable, all 5 variables. For us I have written 5 variables, for you maybe 10 variables, you simply enter, name of those 10 variables.

As per given in that particular document, you cannot just name on your own. Whatever is given in the original document with that name, you enter that name here using then file name of that particular name which we are going to extract with a comma you need to enter, because comma will clear the earlier operation in that Stata if any is there.

(Refer Slide Time: 07:25)



The infile command always require something that is important dot dct file, the dictionary file. If it is a free format, then infile requires the dot dct file for sure. File name with dot dct and infile entry is there and with this fixed space is there. When that space is there, then you apply this particular extraction command. One of the note here is that the using modifier in this case requires the data dictionary file that is in dot dct, not a raw data file. Both the syntax will give the same result.

Like we have already mentioned here, you know the file name and you know the command name, that is in dot dct, this approach as well as the next page approach with the dictionary file that will give you the same result. I am not guiding you much because this type of data, NSS, even IHDS, even NFHS we are guiding is not available. So, we will guide you such type of data which we are going to use it for our purpose at large.

(Refer Slide Time: 08:36)



So, let us come to delimited format. This is even better, a better format than compared to the previous one. We are going to tell you the advantages and disadvantages of each of the format. First one is free format, we told you. Delimited ASCII text format uses other delimiters than space that is very important to note. So, earlier case it is only space, but other delimiters are important. What are the delimiters here? The tab-delimited so you should have a tab-delimited file also or comma-delimiter. So, comma-delimiter, I will tell you what kind of tab-delimited and comma-delimited we are going to guide you.

So, comma-delimited are most common. These two are most common format in this set up, but any special character such as: at, hash, dollar, power and or asterisk also can be a delimiter, but the most common are tab. Just simply tab then another variable or comma then another variable, no need to have space. The most common form is CSV file that is also called comma separated file, usually it is in comma format, where the delimiter is a comma, but the delimiter could be a space, a tab or in theory just about any other character or set of characters, any other. So, we are broadly going to talk about CSV or tab-delimited approach.

The extension of CSV file is written like dot csv. I think in our data I will guide you. This is our data. look at this CSV file. if I just right click here, we need to right click on this, so properties, so it will show you here at the end, it gives you also some information, otherwise, .csv information generally comes here .csv. Here you get the dot csv information. Otherwise, it also written if you enter the extension file here in your size, type, data and it will also show you in one of the column. If you take all those options, it also shows you the CSV option. There are many approaches to understand whether it is in CSV or in any other one.

Let me proceed. So, what I will do, I will guide you step by step. CSV is able to deal with complicated and ill-organized data from spreadsheet and data sheet. Generally, spreadsheet cannot organize everything. Therefore, this data compress as well as organize. Save with limited space and give you a better format and for analysis. So, all the big, large scale data adopts these techniques.

(Refer Slide Time: 11:53)

This format often has a list of variable names at the first line.It looks like
ENTID,State,b2_q204,b2_q210,b2_q211,b2_q212,b2_q213,b2_q216,b2_q218,b2_q219 6950911502,9,1,6,1,2005,1,2,2,2 8383911101,3,1,2,2,2013,1,1,1,1 5116511403,21,2,1,1,2013,1,2,2,2
 If your data is in delimited format, you can use the insheet/import command. insheet using filename, clear import delimited filename, clear Or you can also import it from file menu File -> Import -> text data (delimited, *.csv,) -> browse ->
စို ့ swapan ဖြ

This format often has a list of variable names at the first line. Most importantly you need to understand that this, it looks like this. First line the first row generally contains the variable names. Here in our data we will show you the enterprise ID, state, other variable, these are variable names from the beginning. Then other rows are the observations or the cases.

(Refer Slide Time: 12:24)



I can show you that, so it is here. I will go to our file we can open it, the CSV file. This is the CSV file. But we have opened in Stata, we need to open it, check it with right click, open with at

the top, open with notepad. You look at the data which the sample is given to you, the example is given to you. It is entID, state, then all the variables from the beginning and all other variables are separated by comma. You look at very carefully all are separated by comma. all the observations we have already shortlisted are given here. All the observations are given for our use.

Let me proceed further. So, what I have mentioned here to you that we have already discussed you this. So, this is the one I have already clarified. If your data is in delimited format, you can use the insheet command or import command. So, the command is given here. Insheet using file name then you can able to extract. I will show you right now. Insheet using that file name and then that will extract the data. Otherwise, import delimited file and clear. If you do that, you can able to extract it. Let me just have a check from our data.

(Refer Slide Time: 14:20)

	Computer	Removable Disk (G)			• •	Search Removable Disk (G)
EN	Organize • 🚯 Open	Print Burn New folder				# · 🗋 0
69	🚖 Favorites	Name	Date modified	Type	Size	
0.	Desktop	Reports	7/13/2020 10:30 PM	File folder		
83	🔒 Downloads	Stata15	6/18/2020 5:58 PM	File folder		
51	💫 Recent Places	State_wise_Data_IN_ASCII_FORMAT	7/13/2020 10:30 PM	File folder		
		Supplementary_DOCUMENTS	7/13/2020 10:31 PM	File folder		
D 16	💢 Libraries	EC6A_ST05_UTTARAKHAND.dat	6/8/2016 5:32 AM	DAT File	22.327 KB	
L It vol	Bocuments	EC6_Dictionary.dct	8/17/2020 12:31 A	DCT File	1 K8	
, •	💰 Masic	my_do.do	8/4/2020 2:25 PM	D0 file	2 KB	
com	S. Pictures	EC6_uttarakhand_extracted.dta	8/17/2020 12:34 A	DTA File	21,573 KB	
00111	S. Videos	Example data_set.dta	8/6/2020 3:18 AM	DTA File	17,754 KB	
inche		practice_dataset_IHDS.dta	8/10/2020 5:01 AM	DTA File	48.152 KB	
1112116	Computer	practice_dataset_NSS73_block2.dta	₽8/6/2020 2:26 PM	DTA File	19,009 KB	
	 Local Disk (C) Local Disk (D) 	Type: DTA File	8/17/2020 2:32 AM	Microsoft Office E.	2,388 KB	
impo	 Local Disk (E) 	Size 18.5 MB	8/17/2020 2:40 AM	Microsoft Office E	3,351 KB	
		Lecture 20- Val Date mobilied and table in that	8/21/20/0 1:58 PM	Microsoft Unice P	1,587 8.8	
Orvoud		Decture 21: Extraction of unit level data in stata	9/21/2020 3:42 PM	TAR GIA	1.997 ND	
OI you (K Network	14331-160-080				
e-1						
File ->						
1.1.1.1						
cubmit /						

The data is in, so we need to open in Stata, but now what I do I will open the Stata for you. We have to go by the command. This command if you enter, but you require, most importantly your data must be having the file name with csv. If the data is in csv then you can able to extract it. But if the data is not in csv or data look at, the original data is not in csv, it is in that file, the original data.

In fix command like this you can open. So, like this one is our data. If I just, the csv it is there, if I use that extension you can able to command. we have converted it as per our use. Originally the data is not, if you just go to the NSS or even economy sensors, you would not get the data in CSV format. You have to convert it with by file, save as in CSV you can able to convert it to CSV format if you wanted to test whether it is getting extracted or not. So, we have converted for our use.

So, our data here, now you can see it is, I have already shown you just couple of minutes back that it is NSS1 hyphen csv we have named it. I need the path of that. This is in G file; I have to use that for my command for a better extraction. This is, what I will do, I will copy this or I will type it insheet using then file name this I have already just shown you. Now what I will do, I will, where is that command, it is here.

(Refer Slide Time: 16:24)



So, this is insheet, you just check very carefully that insheet, you have to mention very clearly, do not miss any single word, using then file name. So, how to get the file name, file name is, I have already shown you G, it is in G path. So, it is in, it is here, so it is in G path. So, let me copy that first or what you do, just click here or right click and copy it. You just take that to your Stata. You using file name right click and paste it. So, this is your path has come. What else you require, you require the file name also.

So, what is the file name, file name is here, just copy that file name. So, right click and copy. So, I have already copied. So, what I do, I will try to paste it here. So, paste. The extension should be given. So, since it is in dot csv, so csv should be there. There are some original data already given, so you need to enter comma and clear. If you do not clear then there will be some overlapping issues, there will be some problem comma and clear. Type clear, if you click it clear then enter. You will get the information. Now your data is ready. This has been extracted.

(Refer Slide Time: 18:20)

	N								
entid[1]	6950911502								
entid state i	82_0204 82_0210	b2_q211 b2_q212	2 62_0213	b2_q214 b2_q	1210 82_0219			* Variables	
6.951e+09 9	1 6	1 20	15 1	2	2 2			Elter yana	
8.384e+09 3	1 2	2 20	1 1	1	1 1			R Name	Label
5.117e+09 21	2 1	1 20	13 1	2	2 2			entid.	ENTID
6.778e+09 24	7 2	2 20	10 1	1	1 1			S state	State
6.769e+09 24	2 1	1 20	12 1	1	2 2			₽ b2,q204	
1.011++09 27		1 20			2 2			₽ b2_q210	
6.042e+09 15	1 3	1 20	11 1	2	2 2			₽ b2_q211	
5.355e+09 6	1 1	1 20	1 10	2	2 2			₽ b2_q212	
5.912e+09 9	1 2	1 20	1 1	1	2 2			₽ b2_q213	
5.964e+09 9	1 1	1 20	1 1	2	2 2			₽ b2_q216	
9.604e+09 32	1 5	1 20	2 1	2	2 2				
8.194e+09 19	2 2	2 20	1 1	1	1 1			[6] 01 ⁻ 0118	
9.077e+09 14	1 1	1 20	14 1	2	2 2				
9.531e+09 33	1 1	1 20	99 1	2	2 2				
9.519e+09 34	1 1	2 20	p7 1	2	2 2			Variables Sna	pshots
5.356e+09 6	1 3	1 20	1	2	2 2			Propertier	
6.967++09 8		1 19		,	1 1			4 Variables	
5.733e+09 24	2 1	1 20		,	1 1			Name	ettid
5.736e+09 24	1 2	2 20	15 1	2	2 2			Label	ENTID
6.094e+09 11	1 1	1 19	1 1	2	2 2			Type	double
6.199e+09 19	1 2	2 20	12 1	2	2 2			Format	%10.0g
6.300e+09 3	1 2	1 20	12 1	2	2 2			Value label	
8.969e+09 9	1 1	2 20	10 1	2	2 2			Notes	
9.194e+09 19	1 0	1 20	1 1	2	1 1			Filename	
9.059e+09 16	1 1	1 20	29 1	2	1 1			Label	
6.6574+09 28	1 1	1 20	1		1 1			Notes	
6.701+09 27	1 2	2 19	1 1	*	1 2			Vanables	
6.953e+09 9	1 2	1 20	15 1	2	2 2			Observations	72,528
1. M.A								+ Size	1.38M
								Memory	64M
£151									
SE 15.1 Edit Data Graphics St	atistics User Wir	dow Help							
SE 15.1 Edit Data Graphics St Iem Ctrl+O ve Ctrl+S	atistics User Wir	dow Help		1 angle					-
SE 15.1 Edit Data Graphics St pen Ctrl+O we Ctrl+S we as Ctrl+Shift+S	atistics User Wir S S C ·	idow Help D 2	9,401	3 servio	es			Variables	-
SE IS.1 Edit Data Graphics St Iven. Ctrl+O ve Ctrl+S ve as Ctrl+Shith+S tw	abstrics User Wir Market Constraints X I abbel dar	dow Help	9,401 #P	3 servio	163			Variables Triter varia	bles here
SE IS.1 Edit Data Graphics St ien. Cirt+O we Cirt+S we as Cirt+Shit+S tw b	x . label dr	dow Help 2 op activity_gro	9,401 ap	3 servio	es			 Variables ★ Filer variation Name 	Dies here Label
SE 15.1 Edit Data Graphics SR wen	anstes User Wer	dow Help 2 op activity_group	9,401 IP	3 servio	÷s			 Variables	Dies here Label ENTID State
SE 151 Edit Data Graphics Sr pen. Crti+O ve Crti+S ew. Crti+Shit+S ew ename. ange working directory.	anstics User Wer	dow Help 2 op activity_grov ity_group	9,401 IP	3 servio	es			 Variables Fiter varia Name entid state bh e 004 	bles here Label ENTID State
SE 151 SE 151 March Data Graphics St en, Chrl-O we AL, Chrl-Shilt-S two sea. Chrl-Shilt-S two mame. ange working directory. g	abshcs User Wer X . label dr . ta activ activity_g ow	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 ap Percent	3 servio Cun.	e3			 Variables ▲ The varia Name entid state b2,q204 b2,q204 	Label ENTID State
SE ISI Ent Data Graphics St een. Criti-O ve Criti-Sitti-S me. b. ename. ange working directory. g y y	atotics User Wir X Label dr . ta activity_g own Event secard	dow Help 2 cop activity_group r p Freq. we (* vict vice)	9,401 #P Percent 51	3 servio Cun. 28.51				 Variables Titer and Name endi state b2,q204 b2,q210 b2,q211 	bles here Label ENTID State
SE 15.1 Ent Data Graphics St eve Crit-So eve Crit-Solt+S eve	Asistics User Wer X I abel dr. . label dr. . ta activity_g ou Excel ipreads Tret data Get	dow Help 2 op activity_group r p Freq. Vet('s0s's0s) imited 'rex.)	9,401 #P Percent \$1 95	3 servio Cun. 28.55 59.46	e3			Variables	bles here Label ENTID State
SE 15.1 Edit Data Graphica St een	Anstes User Wer Anstes User Wer Anstein Constant Anstein Consta	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 #P Percent \$1 \$5 \$4	3 servic Cum. 28.51 59.66 100.00	es			 Variables Intervana Variables varia varia 	bles here Label ENTID State
SE 13.1 Ent. Data Graphics St ent. Chi-O we Chi-Shith-S we assumed and and the second port port at port the second second second second port port	Interson University of the Werk Werk Werk Werk Werk Werk Werk Wer	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 40 Percent 51 55 54 00nayy 00	3 servic Cun. 28.51 59.46 100.00	e3			 Variables Variables Intervaria Name enid state b2,q204 b2,q212 b2,q212 b2,q215 b2,q215 b2,q216 b2,q21 b2,d2 b2,d2 b2,d2 b2,d2 b2,d2 b2,d	bles here Label ENTID State
SE 13.1 Edit Data Graphers St ene	atstic User Wee atstic User Wee atstick at a finite ter data of file Ten data of	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 40 Percent 51 55 54 00	3 service Cun. 28.51 59.94 100.00	¢3			 Variables frite variat Name endd z0,200 z0,201 z0,212 z0,213 d0,212 d0,213 z0,213 d0,213 d0,213	bles here Label ENTID State
NY 13.1 Ten Data Graphics St enc. Chil-O we Chil-S we as Chil-Shith-S we as Chil-Shith-S we as b b b b b b b	ADDIG Uke Wir ADDIG Uke Wir ADDIG ADDIG X ADDIG ADDIG X ADDIG ADDI	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 up Percent 51 54 00 anut anut	3 servic Cum. 28.51 59.46 100.00 facturing" 2	es			 Variables Variables Variables Inter varia Name entd entd b2,q204 b2,q211 b2,q213 b2,q215 b2,q215 b2,q215 b2,q215	Des here Label ENTID State
SE 13.1 Set 10.22 Graphics 52 enc. Crit-So enc. Crit-Soft-S pon. h. h. g pont. g pont. at . at . at . at . at . at . b. crit-Soft-Soft-Soft-Soft-Soft-Soft-Soft-Sof	Attrict User Were Were Were Were Were Were Were W	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 40 Percent 51 54 00 nour 0)	3 servic Cum. 28.51 59.46 100.00 facturing" 2	es ''teadar' J'service	,*		 Variables Variables Variables Variables Name entid star so so	Label ENTID State
SEESE Sector Data Graphica Se Nen. Cont-So we Cont-So we as. Cont-Sont-so we as. Sector Sector Sector angle working directory. 9	Attrice User Wirk Wirk Wirk Angeleine Verwahl wer Angeleine Schwahl Sc	dow Hep 2 5 5 5 5 5 5 5 5 5 5 5 5	9,401 wp Percent 51 55 54 00ary 00 0)	3 servic Cum. 28.51 100.00 facturing" 2	es	p		 Variables Variables Variables None end rate s2,054 s2,011 s2,	Label ExtTD State
K111 K1 K1 Data Guptica St K1 Data Guptica St K1 CM 55 St St K1 CM 55 St St K1 CM 55 St St K1 St St St St K1 St St <td>Asthick User Wire Asthick User</td> <td>dow kep 2 2 2 2 2 2 2 2 2 2 2 2 2</td> <td>9,401 #P Percent 51 55 54 54 50 00 0)</td> <td>3 servic Cum. 28.51 59.46 100.00 facturing" 2</td> <td>es "trade" J'Hervice</td> <td>r</td> <td></td> <td> Variables ▲ International Nume ■ model ■ model ■ state ■</td> <td>Contract of the second second</td>	Asthick User Wire Asthick User	dow kep 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 #P Percent 51 55 54 54 50 00 0)	3 servic Cum. 28.51 59.46 100.00 facturing" 2	es "trade" J'Hervice	r		 Variables ▲ International Nume ■ model ■ model ■ state ■	Contract of the second
K111 Color Caynes S 61 Data Gaynes S Color 64 Color Color 64 Color Color 64 Color Color 7 Color Color	anthic Use We	dow Hep 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 up Percent 51 54 00 00 00 rcer	3 servic Cun. 28.51 59.46 100.00 facturing" 2 nt Cun	es 	r		Variables Variables Variables None end end store sto	Des here Label ENTID State
Mithil Form Overland Pice Left Disas Overland Pice Left Overland Overland Pice Left Annotation Sectory Pice Left Annotation Sectory Pice Left Annotation Sectory Pice Left Annotation Annotation Sectory Left Annotation Annotation Annotation Left Annotation Annotation Annotation Left Annotation Annotation Annotation	Inthis Use We I about the we I about the sector I about the sec	dow Help 2 pactivity_group ty_group ty_group c Freq. Freq. ed format with a dois ed domat ed format with a dois a dois at data ing 20,68	9,401 wp Fercent 51 55 54 50 50 50 50 50 50 50 50 50 50	3 servic Cun. 28.51 29.54 100.00 facturing" 2 nt Cu 51 28.	43 	,•		 Variables Variables Variables	Label ENTID State
Milli Cell Das Guynes S Cell Das Guynes S Weith Cell Cells Weith Cells Name Cells Name Search Sea	anthos Une We	dow Mep 2 2 3 3 3 3 3 3 4 3 3 4 3 4 3 4 3 4 3 4	9,401 40 Percent 51 54 54 54 00 0 10 10 10 10 10 10 10 10	3 servic Cum. 28.51 59.46 100.00 facturing* 2 nt Cur 51 28. 55 59.	es 			• Vanubles • Vanubles • Vanuble None • end • vanu var	Us here Laber BNTD Sure
KETS	Andrea Une Were andrea Une Were activity_q ou Gereispreadd Ten data on fi Ten data on fi	dow Hep 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 wp Percent 51 54 00 00 cces 5 54 00 00 cces 5 5 5 5 5 5 5 5 5 5 5 5 5	3 servic Cum. 28.51 59.45 100.00 facturing" 2 nt Cu 51 28.1 55 59. 54 100.1	es 	ŗ		Variables	end (MD) Sare (MD)
Milli Left Das Graphics Stress Left Das Graphics Stress with Christophics with Chrit with Chrit	Anthone User Works	dow Hep 2 2 op Activity_group 1 c 7:ceq. ed format 1 red format 4	9,401 Percent 51 95 95 95 95 95 95 95 95 95 95	3 service Cum. 28.51 59.46 100.00 facturing" 2 nt Cu 51 28. 55 59. 55 59. 54 100.0	03 			Vanables Vanables Vanables None Red Rog	Us her Laar END Sale Sale
KENS Construction of the second secon	Anno Che Were Anno Che Were Anno Che Anno Che Anno Anno Che Anno Che Anno Anno Che Anno Anno Anno Che Anno Ann	dow Hep 2 2 2 4 4 4 4 4 4 4 4 4 4 4 4 4	Percent Percent 51 54 55 55 55 55 55 55 55 55 55 55 55 55	3 servic Cum. 28.51 500.00 facturing" 2 nt Cum 51 28. 55 59. 54 100.00	*3 **teda* 3*service n \$1 \$4 60 	ŗ		Variables Varia	ertd (MTD) Sure
KENT Left Das Gupton S Left Das Conford et Christon et Christon e	Annor Use Wee A babel dro t a activity of Control of the activity of the activity of Control of the activity of Control of the activity o	dow Hep 2 2 2 2 2 4 4 4 4 4 4 4 4 4 4 4 4 4	Percent Percent 51 55 55 55 55 55 55 55 55 55 55 55 55	3 servic Cun. 28.51 59.46 100.00 facturing" 2 nt Cu 51 28. 51 51 59. 54 100.0 00 y_group		r		Vandeles	uirs here Label <u>NTID</u> Sale entd (NTID double subby
K (1) Af Das Gaynes M Af Das Carlos W Christones Nature Christones	Anthon Law Work Work Work Angle Constraints of the section of the	dow Help 2 ge activity_group 1/2 Freq. 1/2 Freq. <t< td=""><td>8,401 Percent 55 54 54 54 54 54 54 54 54 54</td><td>3 servic Cun. 28 51 55 54 100 00 facturiog[*] 2 55 28. 55 59. 55 59. 55 59. 55 59. 59. 50 00 75 2000</td><td>es </td><td></td><td></td><td>Vanishes Vanishes Vanishes</td><td>end Children Sule</td></t<>	8,401 Percent 55 54 54 54 54 54 54 54 54 54	3 servic Cun. 28 51 55 54 100 00 facturiog [*] 2 55 28. 55 59. 55 59. 55 59. 55 59. 59. 50 00 75 2000	es 			Vanishes	end Children Sule
A SEE Drag Couples II en Orisio en Orisio	Anthen User WW WW WAR ANTHEN USER WW WW WAR ANTHEN WAR	dow Help 2 2 p freq.	Percent Percent 51 54 54 54 54 54 54 54 54 54 54	3 service Can. Can. 28 53 54 100 00 Can. facturing* 2 53 54 55 28 53 54 54 200 Can. Can. 70 Can. Can. Can. 71 Can. Can. Can.	*3 **trade* 3*service n 51 46 00	ŗ		Vanishes	List here List i NTD State
K (1) Cell Das Gupins S Cell Das Construction We Christone The Christone Network Cell Cell Cell Page Christone Page Cell Cell Cell Page Cell Cell Cell Cell Page Cell Cell Cell Cell Cell Page Cell Cell Cell Cell Cell Cell Page Cell Cell Cell Cell Cell Cell Cell Page Cell Cell Cell Cell Cell Cell Cell Ce	Annes Daw Works	dow Help 2 2 2 2 2 2 2 2 2 2 2 2 2	9,401 Percent 51 54 54 54 54 54 54 54 54 54 54	3 service Cun. 28 51 59 66 100 00 feetvalage 2 feetvalage 2 55 59 55 59 55 59 55 59 55 59 55 59 55 59 50 00 00 00 00 00 00 00 00 00 00 00 00	*3 	r		Vanubles Vanubles Vanubles Vanuble	ertd (MD) Guber (MD) (MD) (MD) (MD) (MD) (MD) (MD) (MD)
MEISS MEISS Next, Carlos Meiss, Carlos Meiss, Carlos Next, Carlos Next, Carlos Meiss, Carlos Meiss, Mei	Anther User W.W. W. W	dow Help Image: Construction of the second se	Percent Percent 51 55 55 55 55 55 55 55 55 55	2 service 24 51 54 51 56 service 100 00 51 28 55 52 56 100 00 00 00 00 00 00 00 00 00 00 00 00		ŗ		Vanishies	ertd (Still) villog
Kill Left Das Guptes 19 Left Das Critico We Critico	Anthon User Work Work Work Work Work Work Work Wor	dow Help 2 2 3 3 4 3 4 3 4 4 4 4 4 4 4 4 4 4 4	8,401 49 Percent 53 54 54 54 54 54 54 54 54 54 54	3 service Cun. 28 51 19 46 10 100 00 00 00 00 00 00 00 00	03 	r		Vandes Vande	Ure here Label DITD State
MEINS MEINS Rev. Construction of Constructions of Constructions construction	Anthes Use WW We want the second seco	dow Help Image: set of the set of	Percent Percent 00 0 0 0 28.4 8 100.0 10 0 0 28.4 9 100.0 10 10 10 10 10 10 10 10 10 10 10 10 10	3 service 28 51 28 52 3 service factoring" 2 nt Cur. 51 28 54 120 00 00 w_group arr	*3 **tade* 3*service =	r		Variables	end Evito State 2010 State State State State State State State State State St
KIII Left Das Guptes 19 Left Das Critico Re Critico	Anthe Use We We want the second secon	dow Help 2 2 3 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	9,401 Percent 13,57 44,57 54 54 54 54 54 54 54 54 54 54	2 service Cun. 28 55 55 66 100 00 fasturing* 2 28 55 55 59 55 50 50 50 50 50 50 50 50 50 50 50 50 50 50 50 5	03 ""trada" 3"service a	1		Vandes	Les fore Laber CMTD Stare
KETS for Day Graphics IP en, Christon en,	Anthes Use WW We We We want the sector of th	dow Help 2 2 2 sp activity_group 2 5 2 7 2 7 2 7 2 7 2 7 2 7 2 7 2 7 2 7 2 7 2 7	9,401 Percent 51 54 54 54 54 54 54 54 54 54 54	2 service 2 can. 2 s 54 3 54 100 00 6 seturing ' 2 5 1 2 s 5 2 5 5 3 2 2 s 5 4 2 2 s 6 2 2 s 7 2 2 5 6 2 2 5 7 2 5 8 2 5	*7 **trada* 3*service n			Variables Variables Variables Variables Variables Variables Variable Variabl	end Units here OMTD Safe Field Safe
M 133 Hen Dus Grayns M Hen Chris Hen Chri Hen Chri Hen Chri Hen Chris Hen Chris	Anthon User WW WW activity activity ac	dow Help Image: state of the stat	Percent Percent 51 54 54 54 54 54 54 54 54 54 54 54 54 54	3 service Cun. 28 51 59 66 100 00 facturiop: 2 51 28. 55 39. 59 39. 50 00 00 00 00 00 00 00 00 00 00 00 00	*3 ***********************************	r		Variables	Lies for Liter for CHD State Profile State 10 10 10 10 10 10 10 10 10 10

T a x		29,4	Import delimited text a					 variables 	
commands here	label drop	activity group	Cite to inneed		1.00			* Filter varial	bles here
mand _,rc			Pile to import.			Browse		Name	Label
G\practice_dataset_N.	. ta activity	_group	Delimiter					entid	ENTID
variable b2_q218 'co	activity_gr		Automatic		Treat sequential delimiter	rs as one		state b2.o204	21916
G\Example data_set.d	oup	Freq.	Use first row for variable	names:	Variable case:			b2_q210	
C_MAUOR	1	20,680	Automatic		Lower •			b2_q211	
tivity_group	3	29,401	Quote binding		Quote stripping			b2_q212 b2_q213	
drop activity, group		22.624	Loose •		Automatic •			b2,q216	
tivity_group	TOTAL	14,020	Floating point precision					b2_q218	
define activity_group	. label defin	<pre># activity_grow</pre>	Use default	•	Set range			b2_q219	
value activity_group a	. ta activity	group	Text encoding:						
et using G:\NSS1-csv.c			Western (ISO Latin 1)	•					
	p	freq.	Preview:					Properties	
	manufacturing	20.680	1					8	
	trade	22,447						 Variables 	
	services	29,401	-					Label	
	Total	72,528						Type	doub
	. label value	activity_group						 Format Value Inheir 	
								Notes	
	. insheet usi	ing G:\NSS1-cav						4 Data	
	(10 vars, 72,	528 obs)						Filename	
								 Notes 	
			00		OK Cancel	Submit		Vanables	
	Command							0bservations	
6 0 🔶	D 💿	ii 🔢 1	3					Sorted by	t) N
1511 To Data Gophics Sanst	IS User Window	r Help	3					Sorted by	to n
151 T Data Gaphics Statist G T + L - M - M T + X	ICS User Window	Y Неір 23,4	of a locat denote that	123		- O X		Somed by	6 b h -
Image: State	Iss User Window	 нер 23,4 	C Import definied text	82				Some by	G D R –
Image: State of the state o	Ics User Window 1 Inter Window Inter Window 1 Inter Window Inter Window	Help 23,4 activity_group	of Inport delimited text e File to import	29		- 00 <u>H</u>		Some by	bles here Label
Image: Second	LS User Window LS User Window	Help 23,4 activity_group _group	of Import definited text File to import	10	(c	, m <u>H</u>		Somed by	Dies here Label ENTID
B31 It Dras Gaphics Statistic	Is: Use Window Is: Window	Help 28,4 activity_group group	C Inport delimited text File to import Open Com	ista rr • Removable Disk (G) •	<u>[</u>	5 (11))) (11)	movatie Data _ P	Somed by	ENTID State
Capter State C	Les Der Widow activity activity_gr oup	v Help 29,1 activity_group group Freq.	Import delimited text File to import Open	ista 11 + Removable Disk (G) + 21	(c	5 回 耳 • 11 Sect A	revention Data · · · · · · · · · · · · · · · · · · ·	 Variables Variables Variables Fite and state b2,q304 b2,q304 	Eles here Label State
Das Gupha Stehn Das Gupha Stehn T Das Gupha Stehn T 2 X Ormatable Das Gupha Stehn Ormatable Das Gupha Stehn Ormatable Das Gupha Stehn Ormatable Das Gupha Stehn Ormatable Or	tes User Windowski tes User Windowski tabel drop ta activity activity cop 1	v Help 29,1 activity_group Freq. 20,640	Inport definited text to File to import Open	iria r • Removable Disk (IG) • er Name	Car mol	• (a) III • (b) Second A	moustie () cir , P • • • • • Size	Somed by ► 2 ● 4 ► Variables ► Inter sand Name entot = 2,024 = 2,024 = 2,024 = 2,024 = 2,024	Eles here Label State
Das Gaphes Sonten Das Gaphes Content Das Gaphes Das Gaphes	two windows	y Help 29,1 activity_group _group Freq. 20,640 22,447 29,461	I Insont delimited text Field b import Open Upprize New Vot Source Location Locatio	er = Removable Disk (G) = er er Nume Reports	Date mod	s (m) 12 • Ity Search A Fed Type 10.10 PM Tipe Lobor	weedet Dat . ₽ Sor	Content by	L R - Label State
Carlos Santa Dara Caphoa Santa Dara Caphoa Santa Dara Caphoa Santa Tara X mana Capactor, daranet, IL, wande bul, 21% co. Subarbo	Our Windowsky	29,1 activity_group /group /req. 20,60 22,47 23,47 24,47 24,47 25,47	Port administration Port administration Port	er = Removable Dia (G) + er Name B Reports Stata15	Der mod 711/202 611/202	• (B) IZ • (4) Search Feel Type Status Status Status Status Status	erresatu Dat. P Ser	Variables	C I I I I I I I I I I I I I I I I I I I
Carlo Capito Series Carlo Capito Ser	Loer Window tics User Window tics User Window tics Coer Wi	N Help 29,4 activity_group Freq. 20,440 22,447 29,401 72,528	2 Topor administration Factoring of Come Com	as • Fencesite Dia (c) + r Ture • Remain • States Dia (C) • States Char, NG • States Char, NG • States Char, NG	Core modi 7/13/029 6(17/02) 7(17/02) 7(• G Z • • • • • • • • • • • • • • • • • • •	examined to Date	Variables ↓ 2 0 4 • 2 0 4 10 </td <td>C N N</td>	C N N
Comparison of the second	tes User Windowsky tes User Windowsky tes User Windowsky tes	v Help 29,4 ectivity_group Freq. 20,647 22,447 23,467 22,461 72,528 • activity_group	Integrate administration Integrate administration Tests temport Tests temport Tests temport Tests temport Tests T	ivi • Ferrosale Dia (G) • • Nare 1. Seatorie: Dia (M) 5. Seatorie: Dia (M) 1. September 2014 1. Septe	СЦ/ОВИТ 7/1/2020 MRMS 1/1/2020 MRMS 1/1/2020	• 1 2 3 • 1 4 Search R Med Tage 15137M For Inform 15137M For Inform 15137M For Inform	ser 527 527 528 529 529 529 529 529 529 529 529	• 2 ● 4 • 2 ● 4 • Variables •	Liber here Liber here Rittlin State
	Label deep - ta activity activity_ef - gep - ta activity - ta	29, 1 29, 1 29, 1 29, 1 20, 60 22, 41 23, 41 24, 41 44, 41, 41 44, 41 44	Incortalinet let f incortalinet let	er + Renceate Dat (2) + er Ranc Banc Banc Banc Banc Banc Banc Banc B	Date modi 7/1/2009 6/1/2009 MeNtS 7/1/2009 4/2/2009	i Gen II i Ia Search II feel Iper Feel Iper Solaw Fee Islander Solaw Fee Islander Solaw Fee Islander Solaw Fee Islander Solaw Fee Islander	sorter the second secon	Vanables • 2 ● 4 • Vanables • Vanables • Vanables • Vanables • Vanables • Vanables • 2,210 • 2,211 • 2,213 • 2,213 • 2,219	Label State
Constant Constan	to: Use Window to: Use Window	29,/1 29,/1 activity_group /group //req. 20,40 22,417 23,401 72,548 * activity_group _group	Import derived term Import derived term Import derived term Import derived term Import	er - Renvester Dati (c) - 1 er Nore - B spont - B spont - B spont - S spont	Car Date modi 7/13/009 CELCOMM 7/13/009 MINIS 7/13/009 MINIS 7/13/009	 Image: Second Sec	resolute (bat , p) • • • • • • • • • • • • • • • • • • •	Variables Variables <t< td=""><td>bies here Labert State</td></t<>	bies here Labert State
Constant Sector Sector Sector Sector Sector Sector Sector Sector Sector Sector Sector Sector	Use Window Use Window Use Window Isbel drop ta activity_gr org 1	* 1909 29, 4 activity_group Freq. 20,460 21,461 23,461 23,461 23,461 23,461 23,461 23,461 23,461 23,538 * activity_group Freq. Freq. Freq. Freq. Freq.	Integrate administration The transmission	ini • Ferenate Dia (G) + • • • • • • • • • • • • •	Date mode 771/0200 641/0200 0541/721/0200 841/171/0200 841/17200		ementar (Dat., P Ser L. 23M KB	Variables Variables Marce Marce Marce Marce State State <td< td=""><td>E III IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII</td></td<>	E III IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
C C C C C C C C C C C C C C C C	Los Une Wordown La activity Cop	v Holp 29,1 activity_group Freq. 20,407 22,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407 23,407	Incontained but I	er + Renceate Data (2) + er Bane Bane Bane Bane Bane Bane Bane Bane	Date mod 7/1/2009 6/1/2009 Methy 7/1/2009 Methy 7/1/2009 Methy 7/1/2009	Im II Im	Ser 2,331 K3	Vanables	Dies here Label State
Comparing a series of the	Los Wedged La activity of activity p manufacture P	V Help 29,1 34tivity_group 7ccq. 20,647 22,647 22,647 22,647 22,647 22,647 72,538 * activity_group group Frcq. 7ccq. 20,647 22,647 24,647 2	Input defined term First tempol Fore For	re Removable Data (c) = 1 re Nume B spaces B spaces	Car Determont 701/0209 6470000 701/0209 40175 4077200	 I a second a seco	2000 (2000) See (2000) (200) (200)	Variables •	Direction No. 1990
Comparing the second seco	Lose Window So L	Hey Preq. 20,4 20,4 20,4 20,4 20,4 20,4 20,40	Inportalimitetari Tarto import Tarto import Tarto import Tarto import Tarto import Torto Tort	ini • Ferenate Dia (G) - • Nare I Feren I Status Status Status Status Nare I Status	Date mode 771/0200 641/0200 MEM 5 771/0200 8471/2200		ementar (Data, P Ser L. 238 49	Variables Variables Marrier Variables V	U Dirs here Liber State
Compared and a series of the series of	Lee Wordow Label deep Label deep Latertivity ge Total Label detinit La activity	x Help 29,1 activity_group Freq. 20,640 22,41 30,640 22,41 20,640 22,41 20,640 22,41 20,640 22,41 20,640 22,41 20,640 22,41 21,41 22,41 21,41 22,41 23,41 24,41 25,41 21,42 22,42 23,41 24,41 25,41 27,52	Importalimity for Importalimity for Importalimity for Importalimity for Import Importalimity for Import	Herework Dat (2) + Ferework Dat	Dar mod Dar mod 41/2020 KLIORMAT 17/12020 KRITZBA K217203	Sound Office State 1 Sound Office	Ser 2.33 G	Vanables VVanables VVVanables VVVVVV VVVVVVVVV VVVVVVVVV	Label Mesheve Label State
Constant and a second sec	Der Windorder Son Der Windorder	* Help 29,1 29,2 activity_group 'group 72,518 a activity_group 'group	Input definited treff Input definit	er Fernevalar Dati (5) + Fernevalar Dati (5	С с	A A	2 23H 43	Variables •	Liber In Meshee Liber State entid Grind State
C C C	Ouer Window Source	Nety 29,1 activity_group	Inor similaria Inor similaria The tampot The tampot T	ini e • Rencoate Dia (G) + e Taran l. Report l. Stati-reg Statistic l. Statistic Statistic l. Statistic Statistic Statistic Report l. Statistic Statist	CELPCMAT 77/1/2020 CELPCMAT 8/17/2020 8/17/2020 8/17/2020			Variables Variables • 2 •	Ders here Labei Bertin State
Comparing a second	Oue Window Inshel drop ta activity activity ge	 ************************************	Importalimite text Far to anyot	es + Renovatir Data (c) + e Parte Bane B	Сал Силе нося 77/13/200 ССС, СОБИЛ 77/13/200 КСТ/2020 КСТ/2020 КСТ/2020 КСТ/2020 КСТ/2020	 Image: Second Sec	Scr 2,3H G 4. 2,3H G	Vanables VVanables VVan	Utes here Label GRTD State
Constraint of the second	Lose Windowski Lose Windowski Lose Windowski Lose Windowski Lose Livity gr L	a thep 29, 4 activity_group recop 70,60 22,447 29,451 72,528 a activity_group 1 20,60 22,447 29,451 72,528 a activity_group 60,4651-cev 29,467 29,451 29,451 29,451 29,451 29,451 29,451 29,451 29,451 20,451 29,451 20,	Insort administration Insort administration There to import	er Fernosik Dat (c) + Nare Baroni (b) Sopon Baroni (b) Sopon Baroni (b) Sopon Sopon (b) Sopon Sopon Sopon (b) Sopon	Сц / Сели 273/228 678/288 771/228 8/37/228 8/37/228	Image: Second Seco	2 2 34 5	Variables Variables Variables Inter units of the second	Label State erind State
Comparison of the second	Lose Window Lose Window Lose Window Lose Viewer L	Nety 29,1 activity_group req. 20,640 22,640 22,640 22,641 22,640 12,528 * activity_group 1 1 20,640 22,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 23,441 24,451 258,451-cav	Inconstructioner File to import File to import Organize New State Sources	in a ferenaliz (bit (c)) e • ferenaliz (bit (c)) e • ferenaliz (bit (c)) ferenaliz (bit (c)) ferenal	CLI COMMT 77/1/2020 CLI COMMT 77/1/2020 8/17/2020 8/17/2020		Ster	Variables	E In
Compare the second	Der Windor ta Obr Windor ta activity activity or ta activity or	Netp Z9, // Activity_group Treq. 70,640 22,441 72,528 * Activity_group Treq. 72,528 * Activity_group Treq. 72,528 * Activity_group Treq. 72,528 * Activity_group Treq. 72,528 * Activity_group 72,528 * Activity_group 72,528 * Activity_group 72,528 * Activity_group * Activity_grou	Ingot Animal binn Fight Ingot Compo Comp Comp Comp Comp Comp Compo	es Personale Data (c) 1 Personale Data (c)	Defe modi 7/13/009 CL (CRMAT 7/13/009 K17/2007 MINIS K17/2007		ала так, « с. 238 са с. 238 са с. 238 са с. 238 са с. 238 са с. 238 са	Varables <	E Raine Contraction Contractio
Comparing a second	Lose Window Solution	a Help 23,1 activity_group '_group Freq. 20,60 22,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 23,41 24,41 25,41 23,41 24,41 25,42 40,411 24,411 25,42 40,112 24,411 25,42 34,411 35,411 36,111 36,111 37,524 38,411 39,411 31,411 31,411 32,411 34,411 35,411 36,111 31,411 31,411 31,411	International transformer I	es • fernosale Dal (c) 1 * Nare 1 Septem 1 September 2014 1	Сц. Сансен ок. Сансен Ок. Сансен Сансен Сансен Сансен Сансен		C 238 G	Variables Varia	Electronic Control Con



So, we can see your data also. This has been extracted. This is a sample for you. This is insheet command we have shown you. Similarly, you can also apply another command that is import delimited file name then clear. The same approach, the same result you are going to get and you can able to extract it.

Another approach is you require either csv data or the delimited data. So, in our file look at this. So, we have, this is our G file. We have the NSS delimited data that is in. So, what I do, I will click and show you how to get. So, you have to go first to the file then import then text it, then you have to carefully observe CSV, if it is in CSV format, then you click on CSV, then submit okay you will get it like this. Then file import then here this is the one text data CSV format.

You click on this. Browse is there. You simply click on browse. You have to go to the data. This is the CSV we have shown you. submit. you need to replace it. Replace command we have not given it. So, do you want to continue and replace the data, yes. You need to replace the previous one then it will replace and you can get it, where the replace one. Since the data we already extracted and it is already available here. Otherwise, you just simply start a fresh document, start the fresh Stata page and go by this file and import then CSV data. It will extract it as per your use.

Now we are going to discuss the detailed information about the other format that is fixed format. So, far we guided you the free format and the delimited format and how you can able to do it, but as I told you, subject to the availability of data. If the data is in CSV or delimited file then you need to apply this command that is insheet or import delimited. Any command you do it, it will extract correctly the way we have shown it. Our next lecture will be on fixed format data and its extraction using the core dataset. With this, let me close here. Thank you.