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

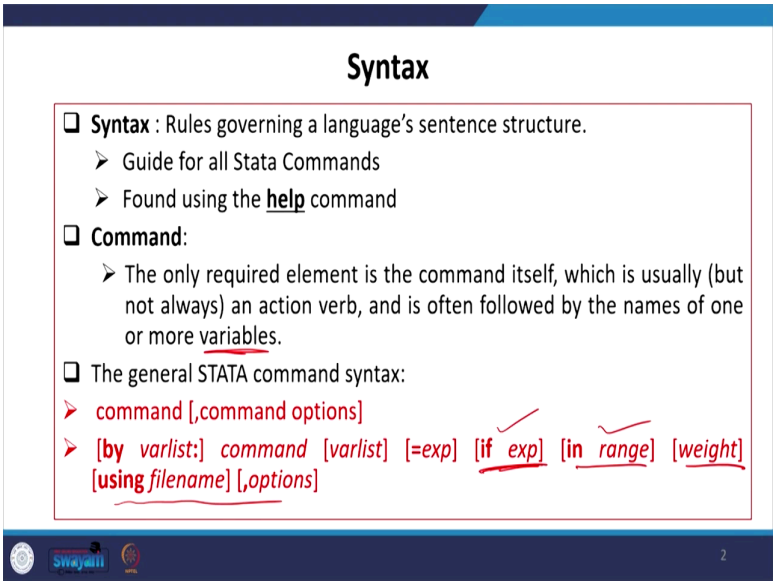
Lecture - 17
Syntax and Do Files

Welcome students to this NPTEL, MOOC module on Handling Healthcare Data. We are in the 4th week to understand the data and the data software, and we will have one lecture on SPSS as well. I hope you might have been following all the previous guidance correctly. Since, I told you repeatedly that we are going to have practical sessions as well, and that practical's will be very helpful for you; for your understanding. Now, in this lecture; we will be clarifying to you about Syntax and Do File in STATA.

So, you have to take the STATA. user access for one week at least; they provide one week; subscription-free subscription-based software to you based on the operation; based on your registration.

So, those aspects we have already guided you. So, let us move on and clarify what do you mean by syntax. The Syntax has basically required some sort of language, some sort of sentences; guidance in STATA commands, that will help us to derive results with the help of syntax.

(Refer Slide Time: 02:11)



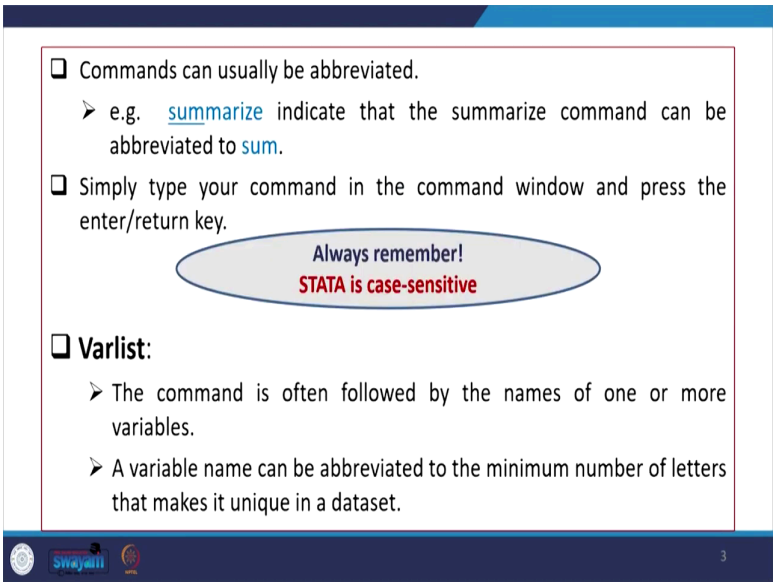
These are called rules governing a language sentence structure; this guides all the STATA commands, and it is found using help command. This syntax, you can actually get it using the help command. On the help command, we have already shown to you, on the main window of STATA, you can just search on the bottom command, you can type help or search your command. So, this is going to give you various syntax. So, what do you mean by command then?

This is the only required element, and is the command itself which is usually an action verb and is often followed by the names of one or more variables. So, it is the action, the exact action you wanted to do, followed by the variables one or more variable. The general STATA commands in syntax los like this; you give your command, and then within the bracket or the end, you have other options; command options, we will explain all those things through our practice.

Similarly, if you want to sort it; through variables, if you want to specify variables further; you can also enter bi variables and colon and further commands, followed by some variable list. And if clauses are there (if any), you can also define them; Their range can be defined, and weight can be assigned. And like some using file and their option could also be added. If you want to get added with the using file, you can also do it.

Commands can usually be abbreviated; like summarize for example, we wanted to summarize the data by their different mean their range, their minimum value, and maximum value; if you want to get all those summaries of the variable; you can abbreviate the summarize. Like you can type the summarize in total; those are those phone, those entries we have made it on this PPT with blue letter; that means, they are the command.

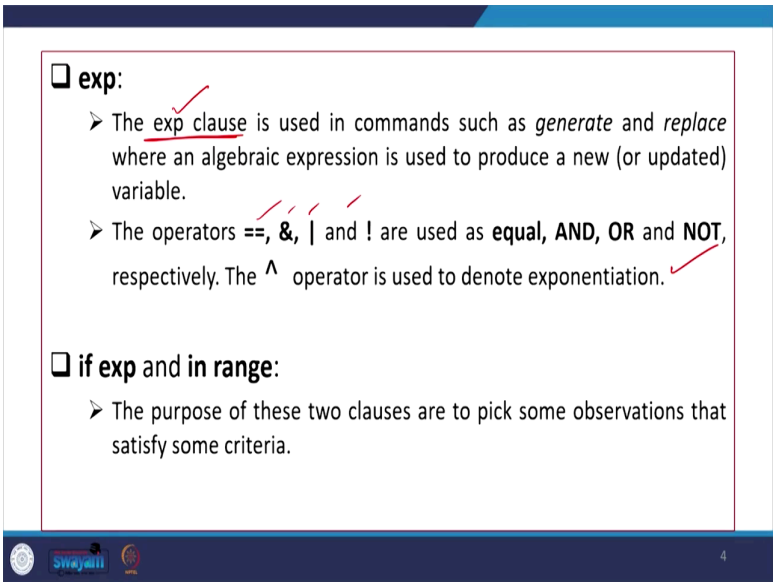
(Refer Slide Time: 04:49)



Command in STATA; so, summarize can be abbreviated to sum; if you simply type sum and the variable name, it will give you the summarized form of the data. Simply, type your command in the command window and press the enter or the return key; that will give you the result. We need to always remember one aspect that, STATA is very case-sensitive.

It is very particular whether you are taking the verb correctly, you are taking the action verb correctly with comma, space, full stop or hyphen. These things are very sensitive to STATA. We have to remember and accordingly follow the steps carefully. Regarding variable lists, the command is often followed by the names of one or more variables. The variable name can be abbreviated to the minimum number of letters that make it unique in a data set. So, as I already said, summarize can be abbreviated to sum.

(Refer Slide Time: 06:05)



The exp you have written that what does this mean. The exp clause is used in commands such as generate or replace, where an algebraic expression is used to produce a new variable. So, the operations such as double equal to used, if you want to specify the value of a variable, you want to get that variable with that particular value; in that case, you have to make the double equal to.

Or sometimes and if you want to join variables and there is a vertical bar that is called or; or on the keyboard, you will find this particular bar that will basically give you options and there are exclamation marks as well, that represents not. So, I think we already mentioned earlier; they are used like, double equal to mention a variable & symbol stands for and; and the vertical small bar is representing or, and exclamation; reverse exclamation mark is basically your, not respectively.

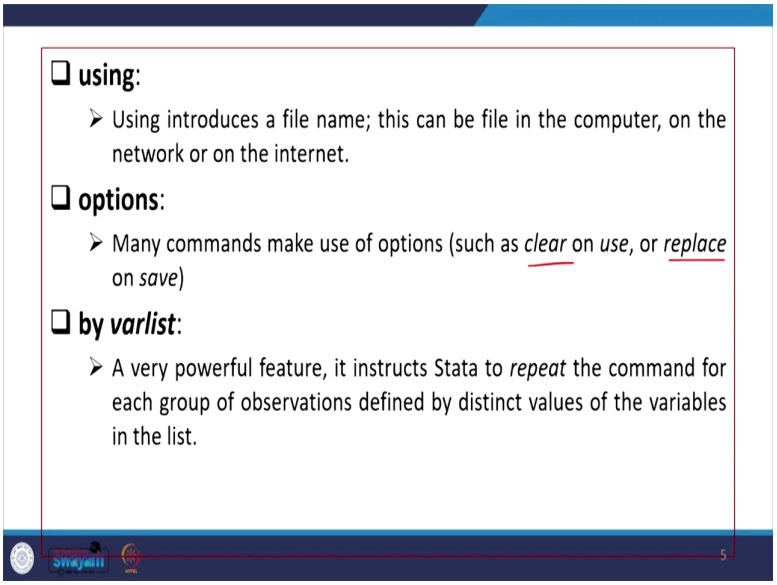
The head mark is basically representing power or exponential exp; exponentiation. So, like x; x square if you wanted to find out a variable square x, cube; if you write down x this, this power notation exponentiation then followed by 2; that will represent x square; in STATA. So, these are quite important, you should take a note of, them very carefully.

If expression basically e; exp clause is basically your expression; it is an expression of what expression you are going to give it. So, it is not the word exp you are writing, rather like here we said if exp we are writing it for expression. What expression exactly, do you want to give it in STATA. So, these are the expression we are in fact specifying. So, exp you are not going

to write it down; had it been the same word we are writing, we could have written it on the command in blue color.

So, this is not the word; rather this is suggesting which kind of expression you wanted to exercise. So, if exp and in range, if such type of aspects is there; if the expression is there or in range expression is required, the purpose of these two clauses are to pick some observation that satisfies some criteria. Like, you want to specify your range of the data; within a limit, so you can specify within a range.

(Refer Slide Time: 09:22)



So, if the expression is important to specify within a range; we will also show it in our practice. Similarly, using then options and by variable list; all those things which we have already listed here, we are explaining one by one; exp or then if expression basically conditioning upon a particular aspect. Then, in range we can define it as a kind of condition, but if you are defining a range a value range. Then accordingly we will be explaining all other aspects using options in STATA.

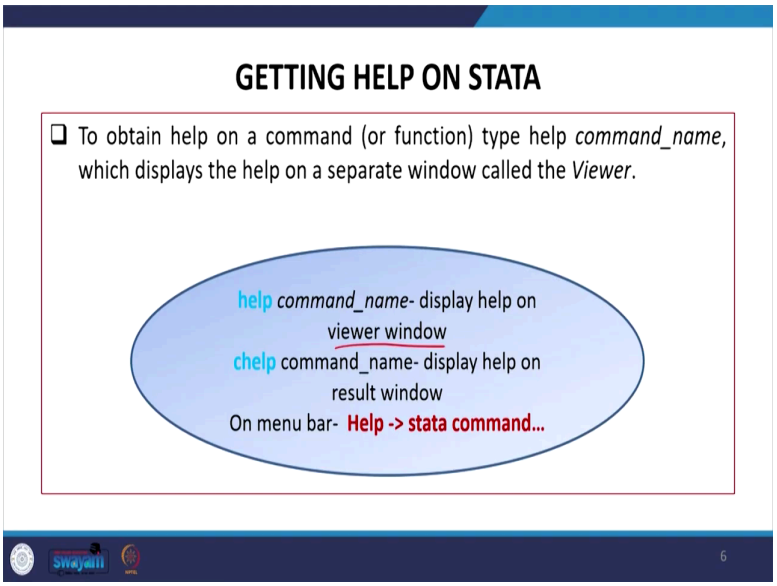
So, in using; using introduces a file name, this can be a file on the computer, on the network, or on the internet. So, if you are using a command that is attached to our command if using as the word is attached; that means, you are actually going to use another file which is stored in either computer or on your desktop or in your disk or on the internet that can be used for the analysis.

Options; many commands make use of options; such as options like clear on the use or replace on save; like at the end, we will be using replace or clear, etc. So, that is going to clear the earlier storage or replace the earlier variable or earlier data. So, this replacement command is going to be useful throughout.

The next aspect is called by varlist; this is a powerful feature; it instructs STATA to repeat the command for each group of observations defined by distinct values of variables in the list. So, it is basically; instructs STATA to repeat the command for each group, of population defined by distinct values of the variable in the list. If there are some distinct variables in the list; then that is going to give us by the variable list, in the specific list you wanted to derive that could be highlighted or derived; so, we will be explaining that.

Now, let us get the help on STATA, how we should go for it and how you can able to obtain it; there is a result with all those commands. The first one is on help, help command or function like you need to for to get that; you need to type command name which; which displays the help on a separate window called the viewer.

(Refer Slide Time: 12:19)

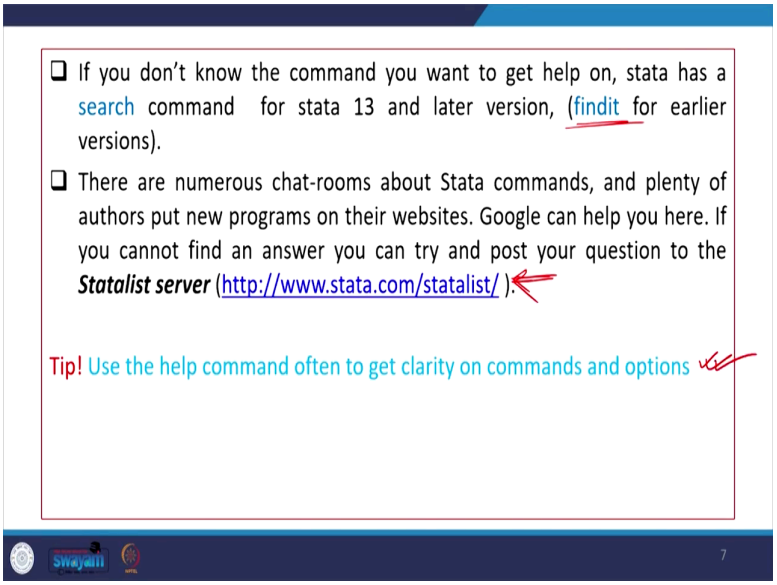


Once too often the help on a command; like or a function, simply type the command name. Command name like command name may be your regression; for example, if you do not know about regression, you simply write e g or regression, on the command bar. This will basically act as; will basically act as a command name or help command.

So, the; once you type that and then click enter, it displays a separate window called viewer; viewer window will come. There is some clarification we will give it here; helps, for example, you write down help, then you come help command them to let it be regression; reg regression complete regression or help let it be about the panel. Or help; you take ANOVA; such type of if you are giving it, it will be displayed in another window that viewer, that is basically called viewer window.

But if you give c help; c help commands, that is going to be displayed on the same result window the; on the same STATA result window, it will be displayed with this c help command. Be you need not open another; it is not going to open another window, rather on the same result window; the help will be displayed. On the menu bar, if you go to help; then click on start command, it will also give you options.

(Refer Slide Time: 14:01)



If you do not know the command, you want to get help on; Stata has a search command for stata 13; in a later version. Find it used to be the search command earlier; in the earlier version basically from stata 13 and before; they used to have find it. You have to find a stay find it that is that used to give the help results. But, these onwards 13 and later versions are giving search command.

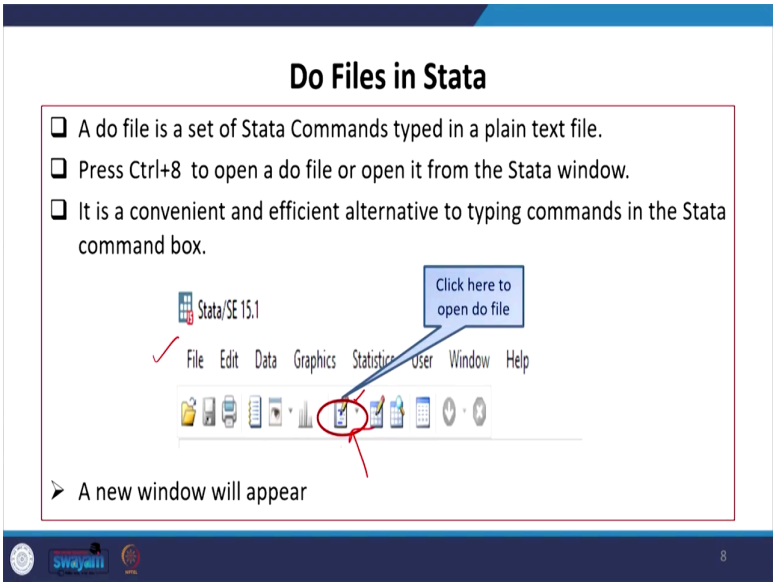
There are numerous chat rooms also available in Stata commands; available in Stata like Stata command. Plenty of authors put their new programs on their websites; then where you are supposed to get those aspects? You need to give; you need to get you may get it through a

Google search. Google helps you; here about, if you simply type in Google about your command and in the state.

Then it will there are lots of authors who have loaded their own calculations, their commands, their do files pliantly available in search engine; in Google search engine as well. If you cannot find an answer, you can try and post your questions to the Stata a list server as well. Still, if you are a bit confused and if you when that; it is difficult for you to go for it. So, Stata list here a list server is in fact, another option to get the help. One tip is there for you; that is use the help command often to get clarity on commands and options.

So, help command is going to be the most and very helpful; it is; it is in fact, quite useful; for those who wanted to be confident with Stata. So, do file; do files in Stata, the what do you mean by that so far, we discuss syntax and its entries, how we can do it. What do you mean by do-file? How in the do file do we enter the syntax? That is important.

(Refer Slide Time: 16:44)



Do files in Stata; on the screen, I will show all those things once again with the direct operation, but at this moment I am just displaying on this PPT, on this slide, it is a snapshot of our window. Like in Stata; 15 versions we are operating on the main tab, then the toolbar tab; I think you will have these options where we will get our do file. If you directly click on the do file, it will automatically open a new do file option.

Otherwise, there is a drop-down; arrow menu, once you click on that it will show whether you will open the existing one or a new one. A do file is a set of Stata commands typed in a plain text file; see it is a kind of text file where all sets of commands are actually typed. Another option is to press control plus 8; control plus 8 to open an and do a file or open it from this Stata window.

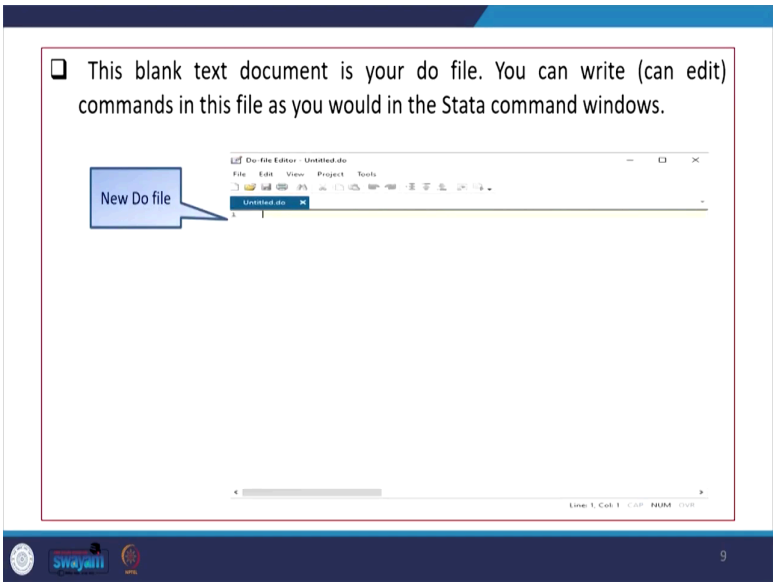
So, control plus 8 is also it is also an important, important direction, an important shortcut to open the do file. So, it is a convenient and efficient alternative to typing commands in the Stata command backbox. Why these are essential, do-files essential? Because there might be lots of commands or syntax, you want to operate for the result or you might have modified your variable, re-coded your variables, like whatever the data is suggesting you maybe in scratch format.

But in those primary data may not; might not have been filtered. You might have a mode you might require further filtration, you might require modification of the variable, or you might follow certain techniques to modify the variable for your use. Maybe a new variable you have generated explains a combination of different other variables. So, if many such important experiments you do with the data, require certain syntax or command; those commands can be stored in do files.

But, if you simply type it and without saving it, without generating a right do file in this operation correctly; on the next time whenever you wanted to operate it, on the next time whenever; you open the data you might not get those files or commands. You have to again type from the beginning till the end which is time-consuming and at the same time; it might be difficult for you to find out the right command with its right byte space.

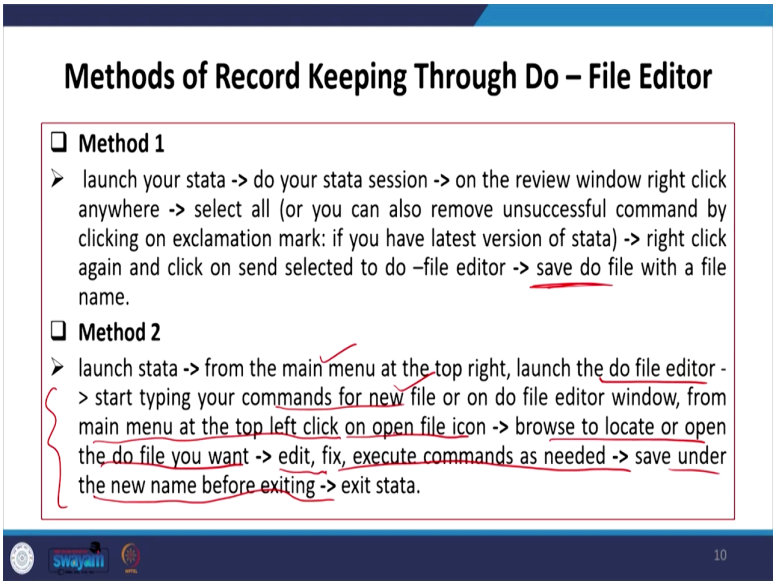
Therefore, it is always suggested that should operate it through the do files. The do-file los like this which I have already told you; you simply need to click on this icon and it will open you do file; a new window will appear and with that, you can able to enter your commands. So, we have been discussing the; do file editor that once you click on the file, you can able to open it.

(Refer Slide Time: 20:41)



This los like the blank page; displayed in this particular slide. The blank text document is in fact, your do-file you can write; that means, you can edit this; on this particular file with the help of your Stata syntax. We will be guiding in through your progress.

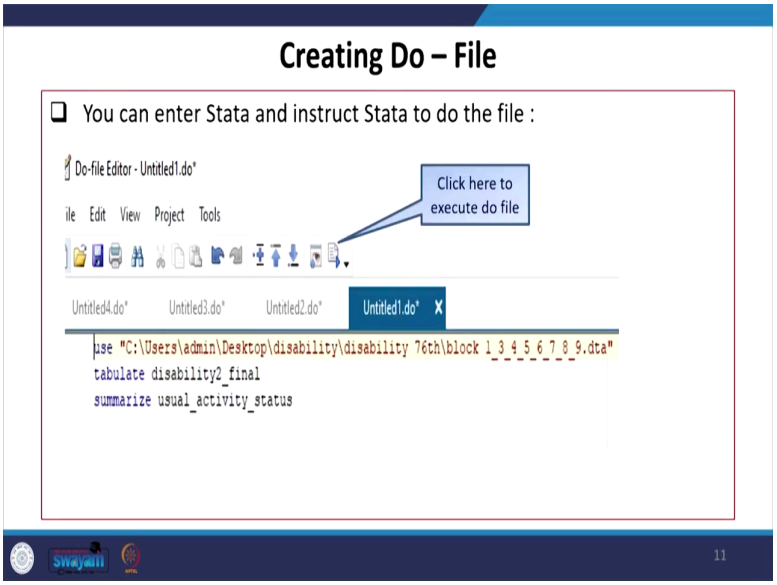
(Refer Slide Time: 21:10)



The, next one is; what are the methods of record-keeping through do file editor. What kind of method you must obtain, must follow to keep your record? Why we are saying record? We say record because; the do file requires a set of commands, different possible commands, and

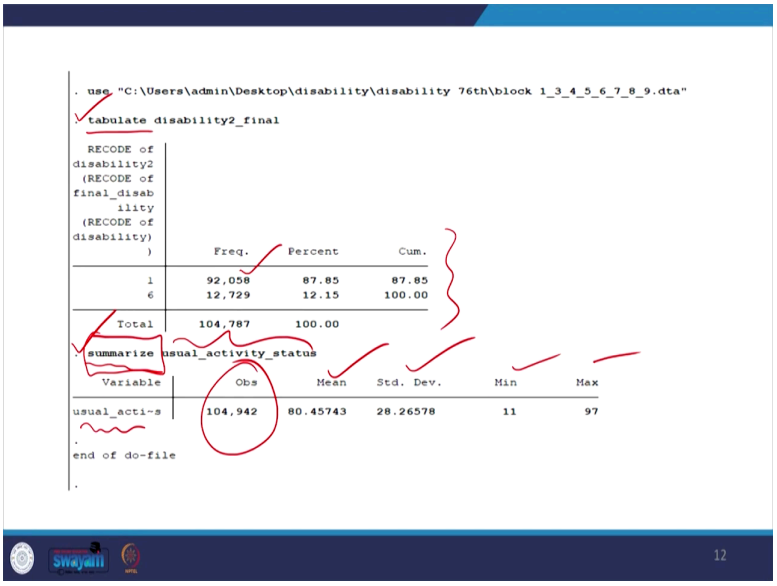
how you should start with. First of all, you launch your stata, you simply start your stata; then I will all guide you in between once again through the stata.

(Refer Slide Time: 21:46)

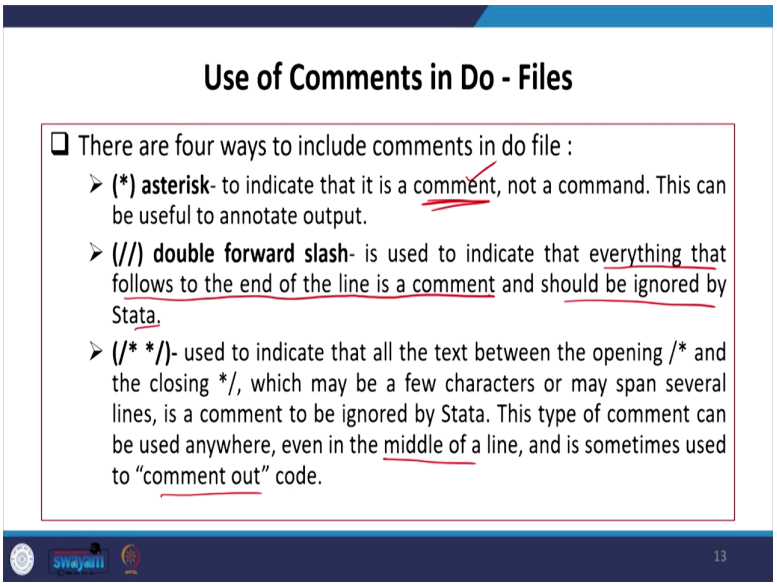


Let me just check once; I will come back to it like this; I will also operate.

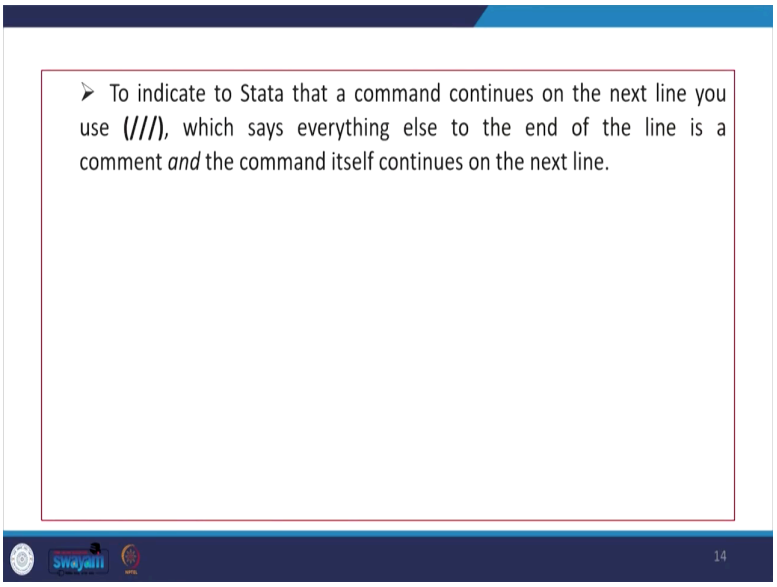
(Refer Slide Time: 21:49)



(Refer Slide Time: 21:50)



(Refer Slide Time: 21:50)

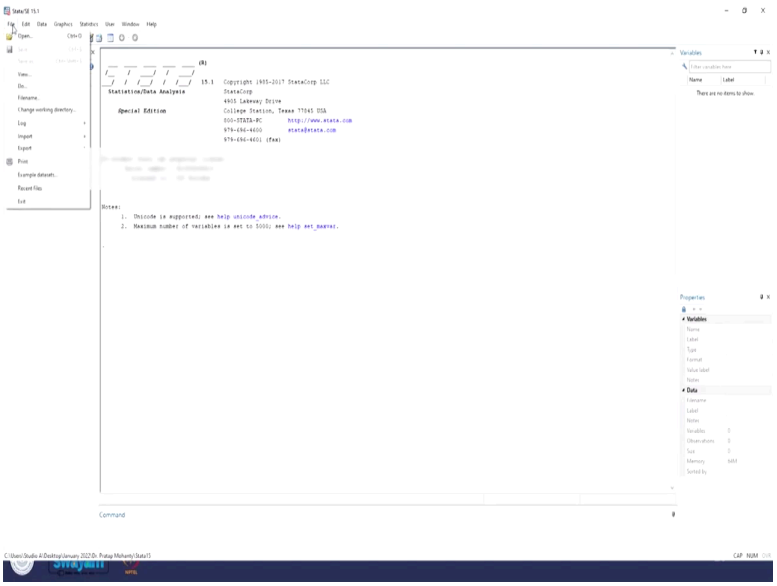


So, let me just guide you first; then I will explain all those things once again. You need to open the Stata, then do your Stata session; then on the review window, simply right click anywhere, select all basically why? Select all is required because you can also remove unsuccessful commands by clicking on the exclamation mark.

If you have the latest version of Stata; that is basically going to give it and then right-click; again and click on send selected to do file, editor, and then save it. That we will operate it, I

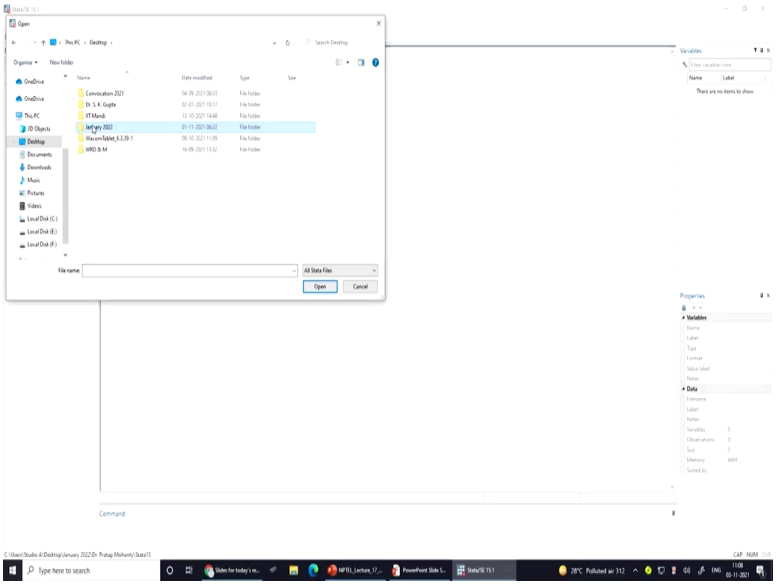
can open the Stata right now; in front of the screen. So, let me just switch it to Stata mode first.

(Refer Slide Time: 22:43)



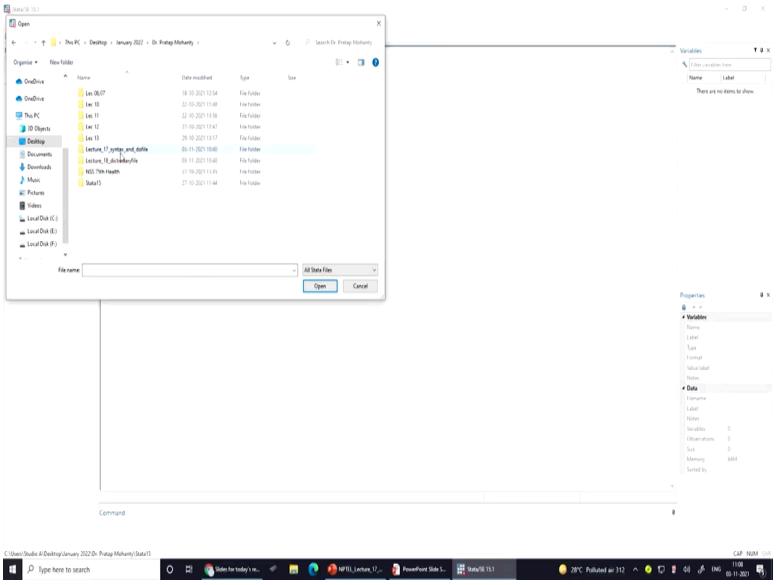
On the STATA, I am just opening one database on your screen. I will let how it works, I am just a sample database; I am opening here.

(Refer Slide Time: 22:55)

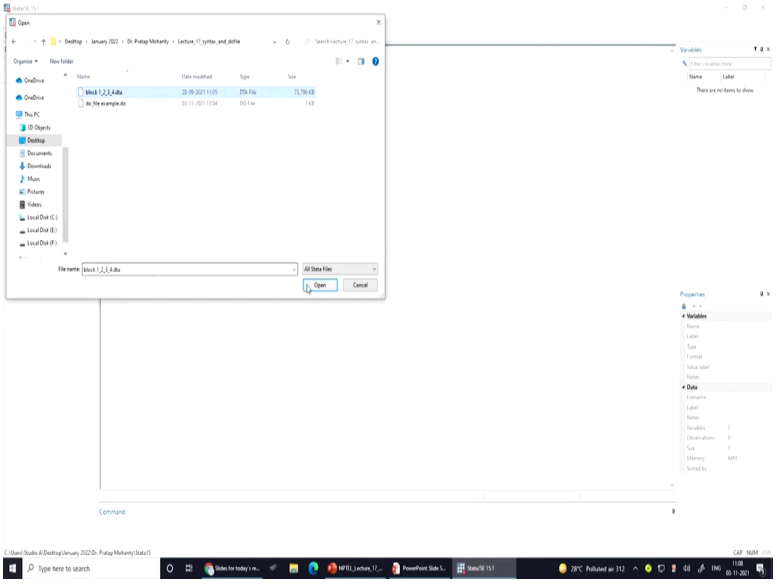


This is our STATA data is being opened.

(Refer Slide Time: 23:00)

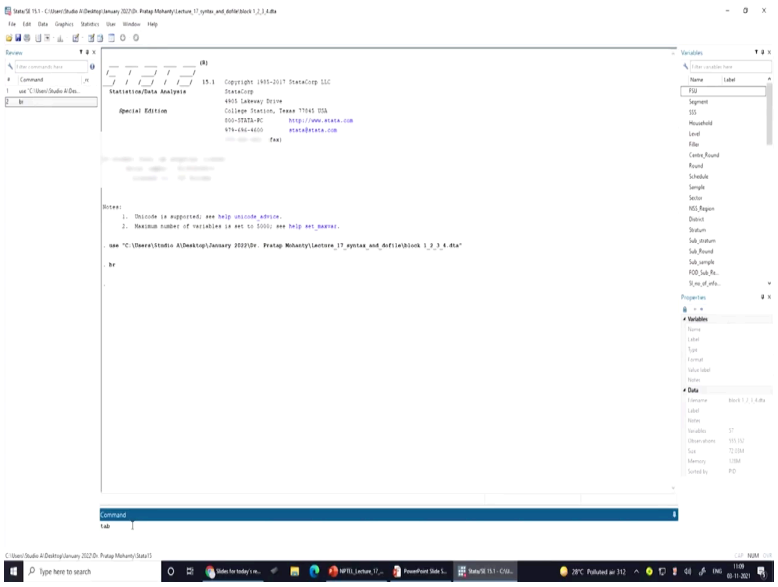


(Refer Slide Time: 23:01)



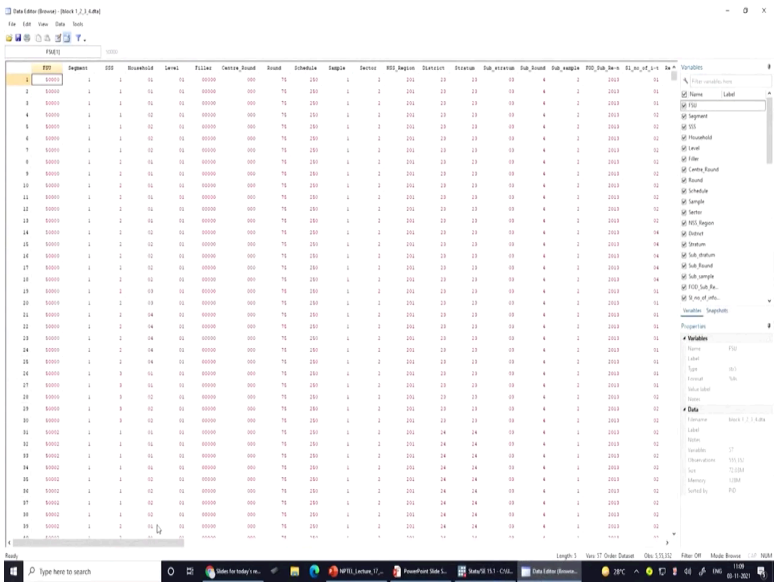
The block number 1, 2, 3 data set; we have opened.

(Refer Slide Time: 23:05)



Now, what I will do? I will operate many things here and how on the review; I will tell you that on the left panel we have a review command. On the review, we can able to filter and copy those things to the do file; I will just do com couple of things. Suppose, this is my data; first what I will do? I will browse it first, here I will browse it; how we can browse it? Simply, will write down b r, then enter it will show us the entire data.

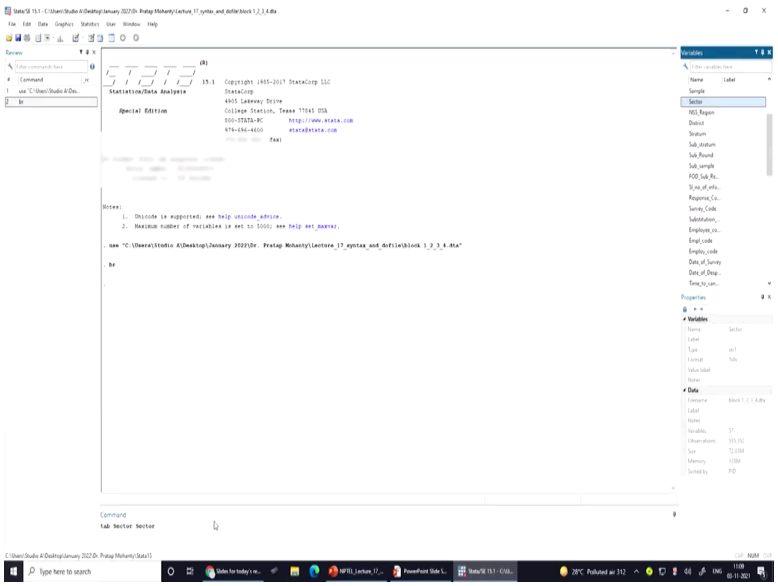
(Refer Slide Time: 23:40)



This is my first command I have given it; then, I can close it and I will do another one. You can just see on the result display b r is written; b r is written. And suppose I write down

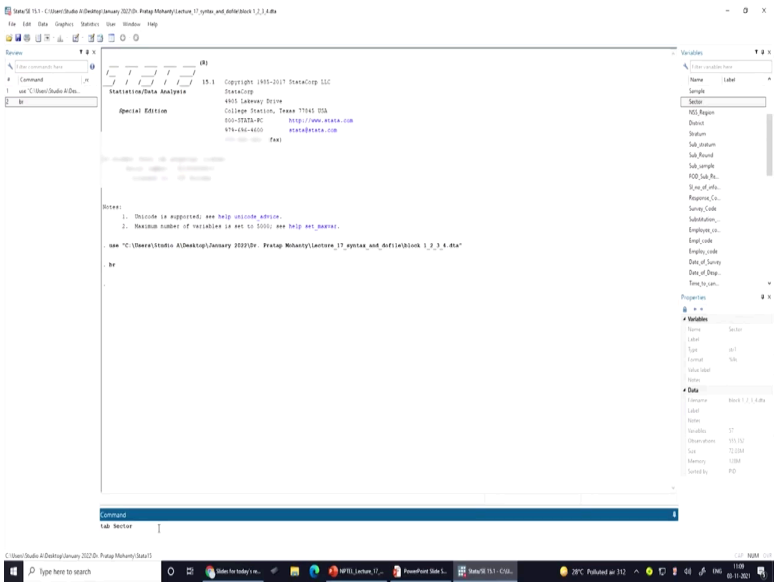
another one like tab, for example, tab t a or t a b or tabulate, I; you will get to know all those things steadily.

(Refer Slide Time: 24:12)

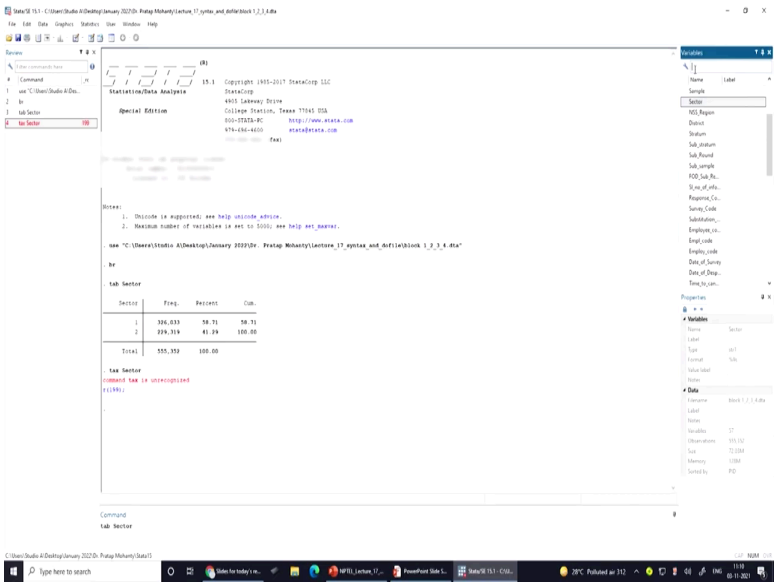


Suppose, I just do tab and I will go a little down and find out one thing from here; that is let it be tab on district, sector let it be on sector; then I will simply enter it. So, this has been entered twice; so I will remove one first then it is entered.

(Refer Slide Time: 24:29)



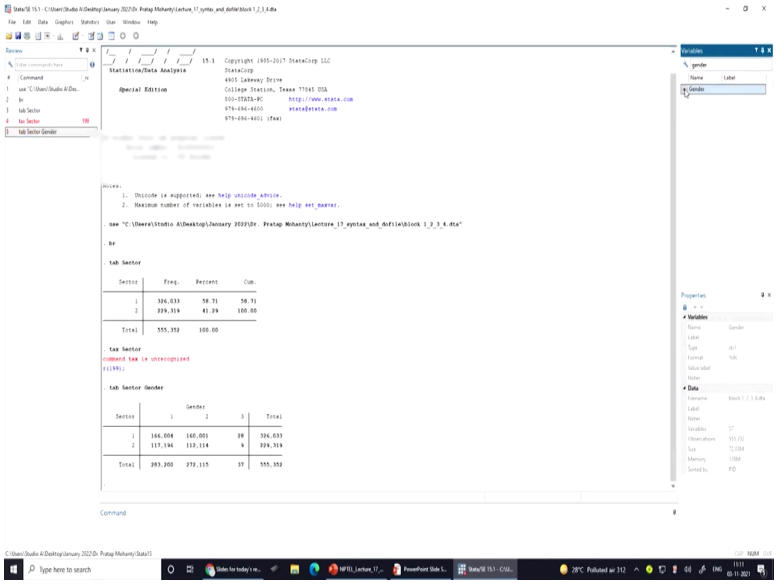
(Refer Slide Time: 24:31)



So, do now by clicking on typing the command as tab or t a or t a b; this is displaying our sector maybe rural and urban. Now, on the review window, you can see that these commands are automatically written. Suppose, I do a wrong entry on my command; let it be tax for example, tax is entered or t a x; tax and that variable on the same valuable that is sector. Show what it comes you just have a lot; now the color is changed, color is now in red; so, it is not recognized by stata.

Suppose, I will again do tab here on the command t a b; tab, then I have entered let it be sector and another valuable two variables let it be; one is a sector, then another is maybe our gender; so, this is gender and enter.

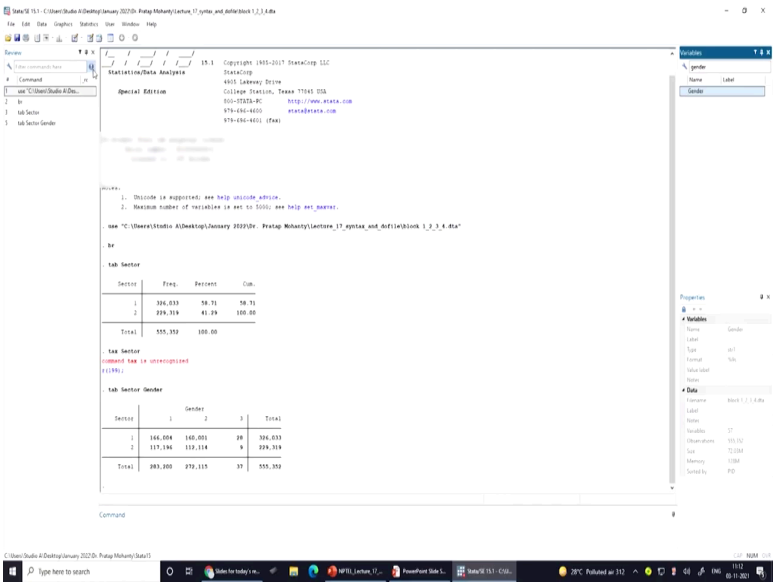
(Refer Slide Time: 25:32)



So, now this since it is a tab is a right command and now we have not just mapped the frequencies, we have also mapped their crossed cross-tabulation. The two variables know their frequency distribution have been together; what other commonalities between these two variables are also identified. Now, in this case what we wanted to do?

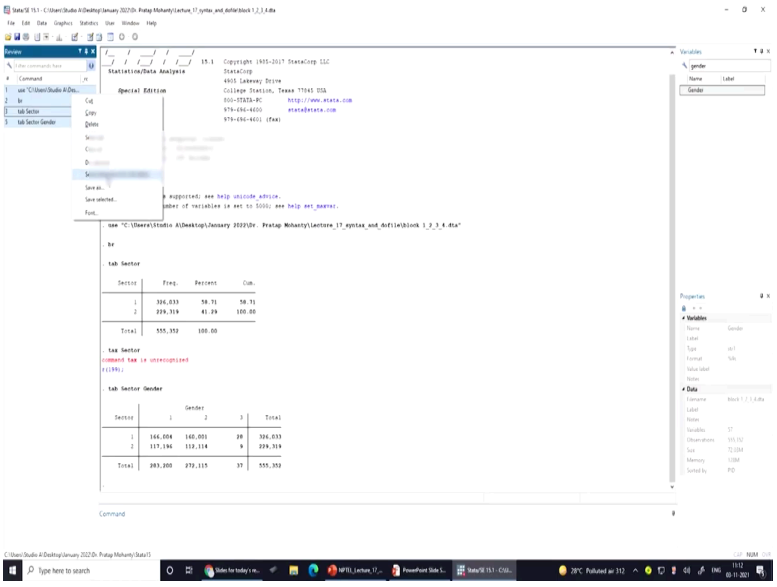
Since, on the review panel, you can just see on the left-hand side on the screen, review panel you have all such commands which we have operated. But, one command that was in red was incorrect, but if you want to keep it all those things in your do file, that is not right. all those since some commands are erroneous; so, better not to take that to your do file. So, we will first avoid the wrong commands, then we take it to the do file that is one option.

(Refer Slide Time: 26:41)



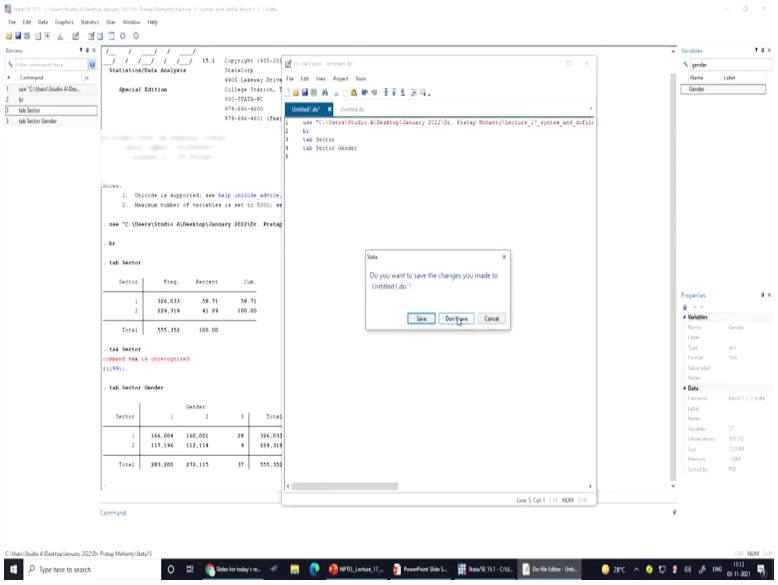
So, we can just do by one thing; I have clicked here, I have clicked here this has actually erased or avoided my wrong entries. Now, I can right-click all though select all I can select everything. So, just select all; So control a or control a or select or control a.

(Refer Slide Time: 27:09)



So, now all have been selected; I will right-click on this.

(Refer Slide Time: 27:17)



Then, I can say select; send selected to do file editor. Now, these have been saved in the do file editor page; the page blank page you have actually saved all your operated correct commands. So, this is one option where you are supposed to follow the do file editing. Now, suppose I do not save it; let me move to the, your PPT once again.

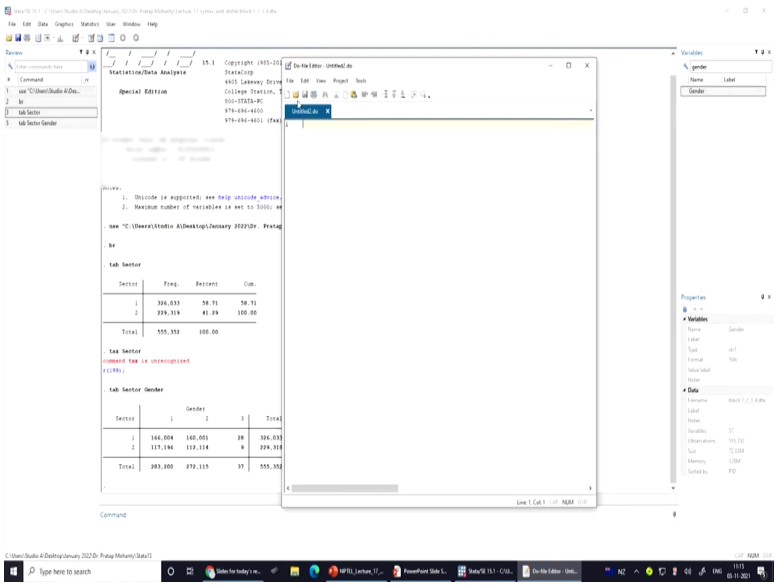
So, this is the one; the first method we have already saved told you, on that do file which was created at the end; if that was my final one, all the commands are fine; then you can save it. So, save we did not do it deliberately; we will follow another method and we will also tell you. Another method, like again we need to go to or you need to open the stata, then we go to the main menu at the top right.

The main menu at the top right; launch the do-file editor; do file editor. Then start typing your commands for the new file or you can file editor window from main menu at the top left click on the open file icon, then browse to locate or open. Basically, you can just follow all those directions; so basically, we will simply open in the main menu, then you will launch the do file editor; then we can start type start typing our commands. Whenever we know that these are my commands and I am quite handy to them; So I can type it.

Then, then what I will do? That I can type it or what I do if my do file editor is already in my window somewhere; I will just get it, click on the main menu or the tape top left click. I mean on the top left; I will click on the open file icon; then browse to locate or open the do file where I have saved it.

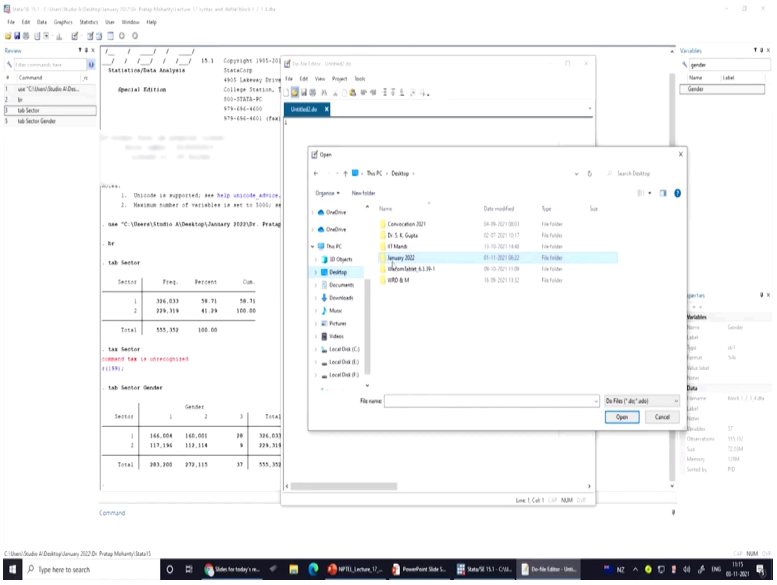
Like I can just open it and let. Then, all those options edit, fix, execute, commands as needed, then save under the new name etcetera can be done it; at the end know we can also keep the exit one as well; this is what I am going to show it to you. So, now, my next aspect is to follow another method for do file here. So, we will; I will first open the new do file window; is not it?

(Refer Slide Time: 30:01)

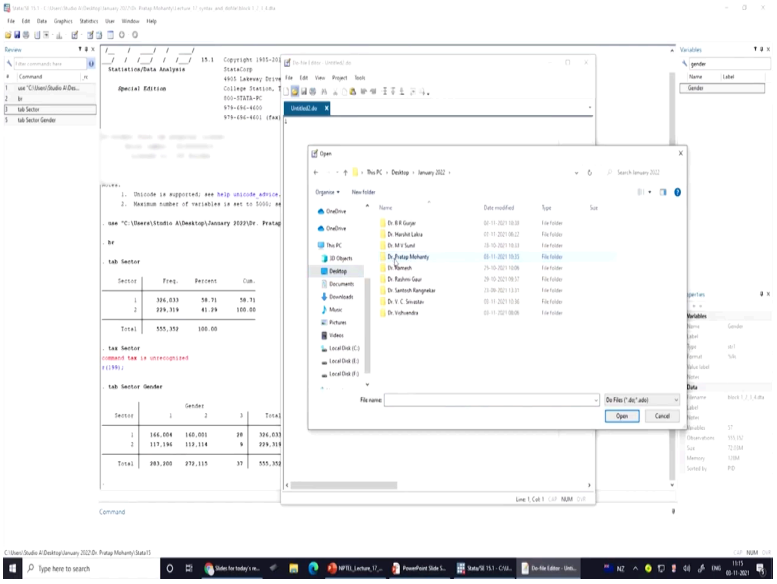


This is where I have clicked it; the do file window has been opened, you can just follow it here on next to the statistics in the toolbar, we have the do file; icon you have clicked it. Now, if you have already saved a do file; then you can click on this folder; it will redirect you to the do file.

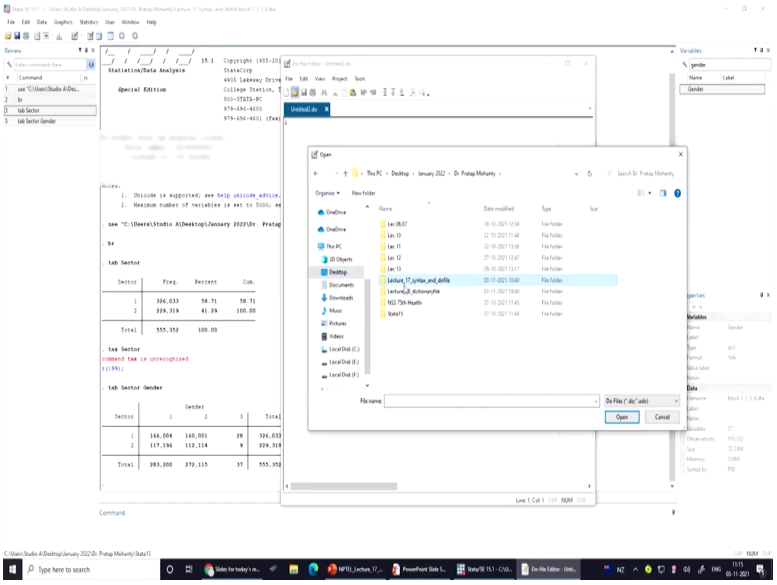
(Refer Slide Time: 30:18)



(Refer Slide Time: 30:23)

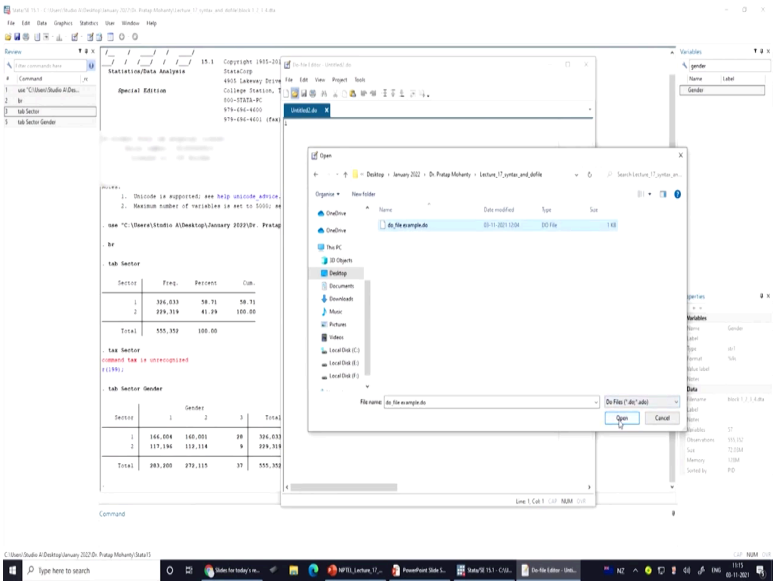


(Refer Slide Time: 30:24)



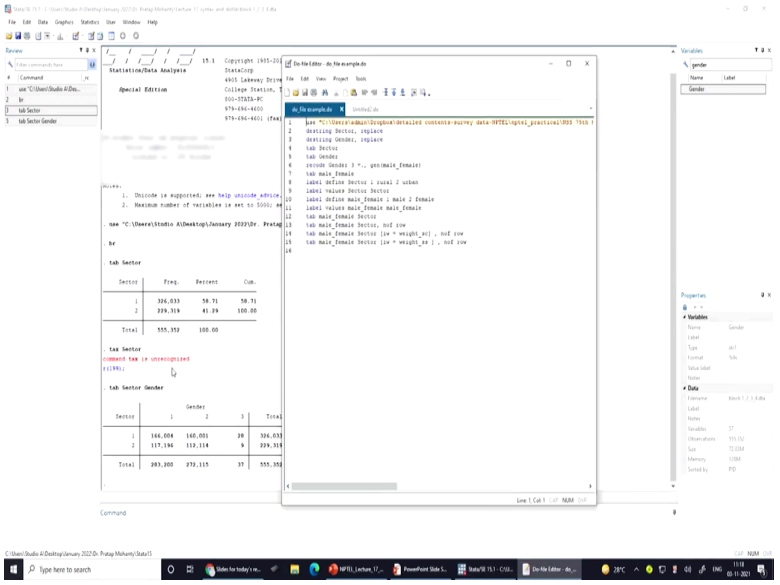
So, the do file we already saved it.

(Refer Slide Time: 30:27)



We have already made; this is an example do file, so we can just open it.

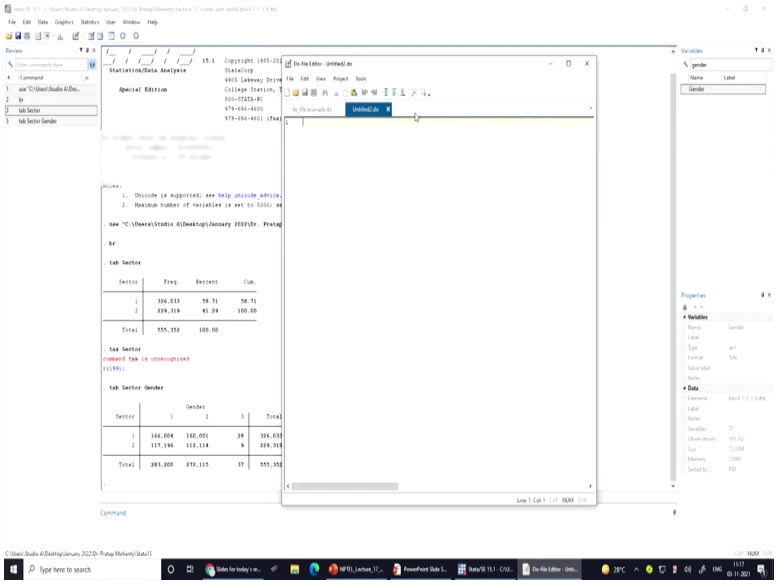
(Refer Slide Time: 30:32)



So, why am I opening it? Just for your reference that know; how you can go for it, how we have operated it, how we have checked different entries, different commands. How those options command is entered, how if the command is entered, how weights are given? So, many things are already mentioned; we will explain some of those things later; So let us first understand.

So, let me just go to the next slide and find out what is new and how it will help you to operate on your own. You can enter Stata and instruct Stata to do the file work for you; right and like here is the one-click here to execute the do file. So, this is the one we have shown on the screen. Now, once it is opened; the you can also open many do-files, like on the next window on the do file one; I can just show it to you here like on title another one, you can also open.

(Refer Slide Time: 31:45)



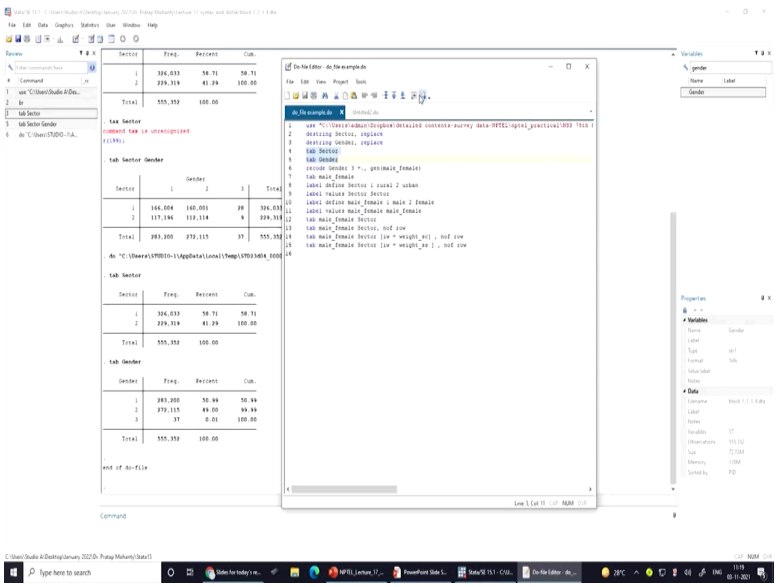
So, some other do file window, you can also simultaneously open. Once you open this, it will also give you another do file to open to the next, but at this moment; it is not required to guide you further. So, let me move to the next guidance.

So, like from the do file; we can execute as I already told you, we can execute, we can summarize. We can draw on from the do file also; like if that was the data, I hope this is the data; like we can just I am opening the screen in front of you and also give you the do file a new window right next to you and we can operate one by one.

Likewise, we manually operated through like we type tab and sector; it displays results, tab two variables; it displays also results instructs. Similarly, here we can also do the same operations.

So, like here I can select these two, let it be tab, sector and gender; I like tab sector, I can execute it; is not it? So, similarly; I if I just execute it; it will actually be drawn correctly and this is the; execute option that you can also click on. So, one just selects this tab sector, you can execute here as well.

(Refer Slide Time: 33:25)



Yes, so by selecting these and executing here; you have got the result as well on the main screen and this has given. Similarly, the entire do file; if you have correctly entered your do file, specified the do file correctly with its correct slash, double slash with right comma, with space.

If you select everything it will simultaneously derive all the results on the screen and it saves your time. So, I am just going to the next aspect. So, we have done those things; summarize what summarize gives? Summarize gives you about information about observation, and summarize of a variable.

So, summarize command we have given it here these are our other variables; this is my command, this is my syntax. And once I enter it; it gives information about the mean of; number of mean deviation, standard deviation, minimum value and maximum value. It also gives from the beginning about; if I click on tabulate and the variable name, then it gives frequency distribution and its cumulative percentages as well.

So, use of commands in do-files; use of commands in do-files, what kind of commands we should make it; we should specify on the do file. There are four ways to include comments in do file. These are very important and quite often used by advance users. Like for example, asterisk mark; if we are entering these indicates a comment; it is in fact a comment. If in the do file, we are given asterisk marks; this is not a command, this is a comment. This can be useful to annotate output.

So, asterisk marks are quite important; this is useful to annotate the output as well. The second option is a double forward slash; not a backward slash, backward slash could have been like this. If it is a double forward slash; it indicates that everything that follows to the end of the line is a comment and should be ignored by Stata.

Like; this indicates everything that follows to the end of the line is a comment and should be ignored by Stata. Like once the double slash is given; this indicates that everything that you that follows to the end of the line; anything that follows to the end of the line is simply a comment; not a command and should be ignored by Stata while running the command.

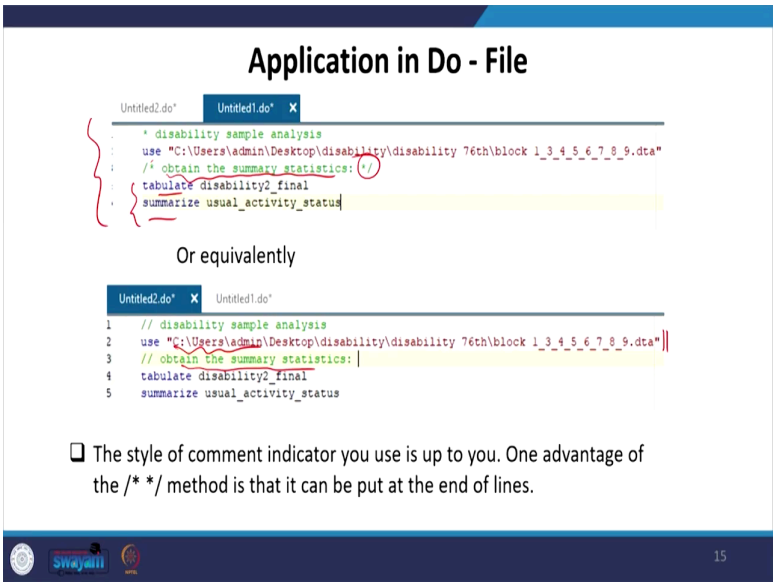
Then, like for one single forward slash with asterisk mark; then we ended it with asterisks marks and then one forward slash; then this is used to indicate that all the text between the opening, that is forward slash with its asterisk marks. And the closing with these which may be a few character characters or may span several lines; that this indicates that this may this; your command might continue to several lines.

So, and it and is a comment to be ignored by; this is also a comment to be ignored by Stata. It is not just comment rather it also suggests, that there are many other comments followed after this particular entry. This type of comment can be used anywhere and even in the middle of a line and is sometimes used to comment out code. So, these are quite interesting and you should take note of it; we will also show it here.

to indicate to Stata that a command continues on the next line; you use once you are saying that your command is actually still continuing like you have already entered a command, there is no space at the end. So, if you want to start fresh on the next line; you should include these three forward slashes. So, these forward lines slash are going to explain to you or read the Stata that you are continuing with the command.

Says that everything else to the end of the line is a comment; everything else to the end of the line, end of the line is a comment, and the com; command itself continues on the next line.

(Refer Slide Time: 38:20)

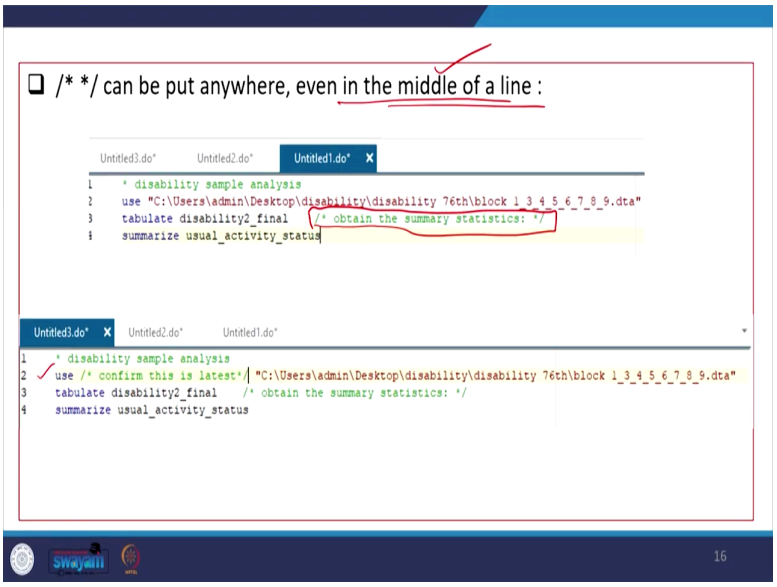


So, like this you just have a lot on the one sample do file, we have shown you that this is your asterisk mark forward slash; one asterisk mark. Then you have given your comment; it is ended with this again. So; that means, Stata is going to read as a comment; not a command, but these are your command; these are your command. Next one is like; so, similarly, if your command you have given basically you have given a command here.

Again, it is continuing that what is this command about basically; it continue for, it ends with the any sort of entries are basically read with a double slash if you have; double forward slash if you have and it then it continues with the com continue with the comments, not the command. So, Stata is simply reading the after this forward slash double forwards slash is also the; the comment.

Alright, the style of comment indicates; the style of comment indicator you use is up to you, how; which style you are actually in including for your command. One advantage of this forward slash with asterisk mark method is that, it can be put at the end of lines. End of the line if you are putting it; that means, it indicates a complete coverage of your command, complete idea of your command.

(Refer Slide Time: 40:04)

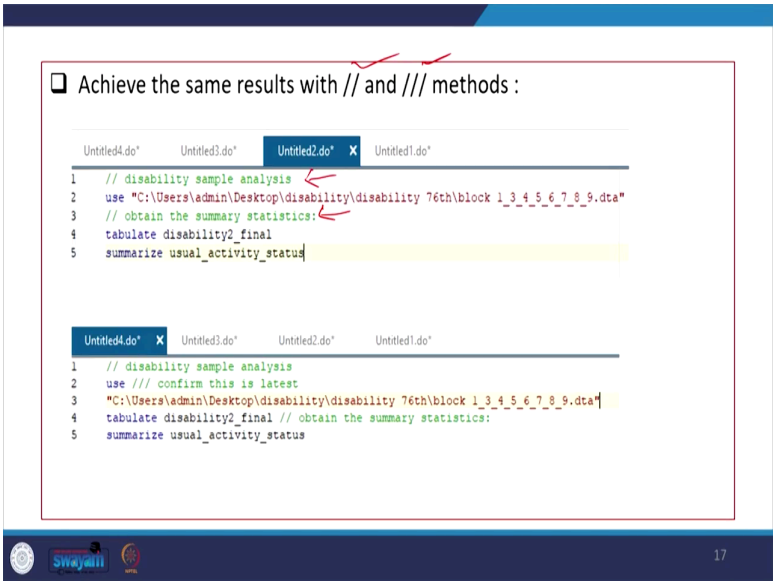


Now, this one can be put anywhere as we have already said; it is dependent on the researcher or on the person who is which is using. Even in the middle of a line; this is what we have already said, middle of a line we can keep it. Like tabulate, suppose you have given then in the middle of a line; you have already we have given this kind of entry.

This indicates that; it is; it is simply a com comment; it is not going to disturb your command. So, Stata if you select everything and run it; Stata is going to read your command, not your comment. Then, similarly in the next example also; we have shown it and like here use for example, here use; star forward slash, then asterisk marks, then confirm this is latest.

Suppose, we have simply given it; this is not going to be read as a command, as simply a comment; so, it is not going to be problematic. And in fact, it is very helpful for the researcher to remember your do file and operate do file correctly.

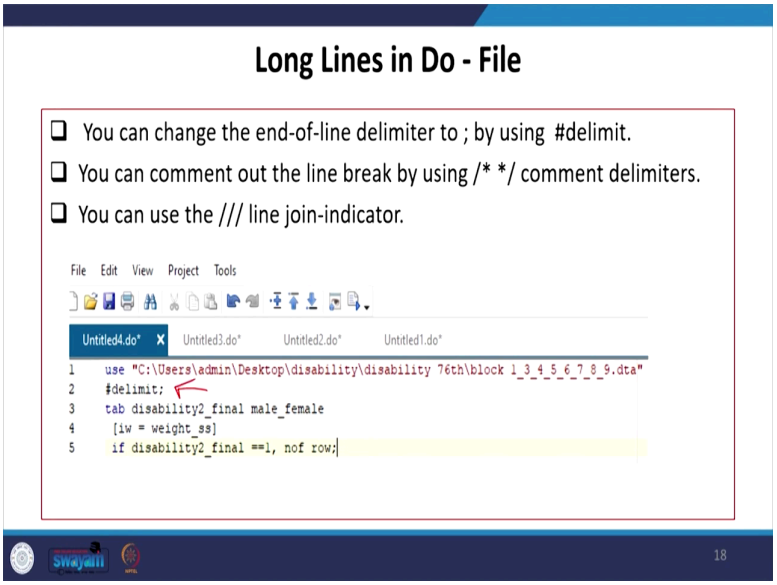
(Refer Slide Time: 41:20)



Here, like double forward slash and triple forward slash; we have already mentioned. So, after whatever it follows for a double forward slash, then that is simply the comment; we have already told you. So, here you can see this; this is simply your comment. Then in the triple; forward slash, it is basically reading as a continuation.

So, continuation and in continuation to the line and usually that is given at the end. For example, you have used this; you have used this triple forward slash. So, the Stata is going to read; as your continuation command.

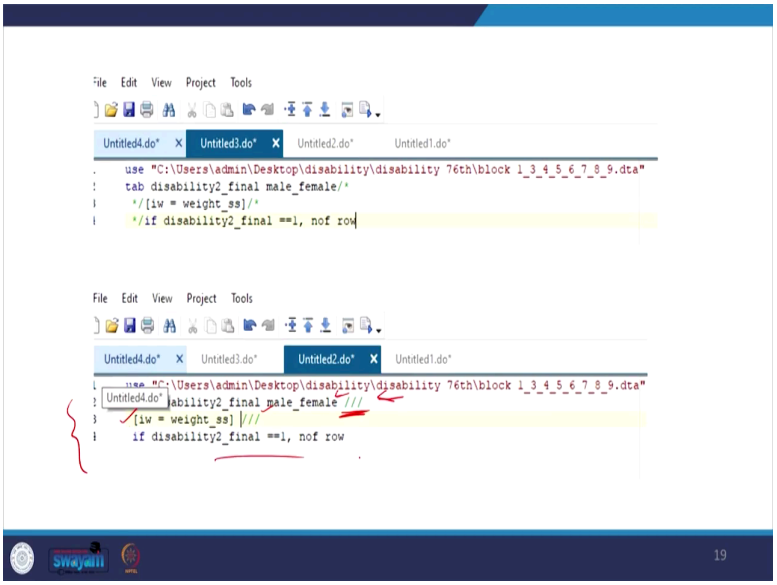
(Refer Slide Time: 42:22)



That means, your commands are still continued; what do you mean by long lines in; how to handle long lines in do file? You can change the end of line delimited to by using delimit. So, there are delimiter line as well; so, you can comment out the line break by using this; with the comment as delimiters.

You can use the triple slash, forward triple slash line to join indicators; this is what we have already guided to you. So, I think already I said delimit you can also use it; as and when required will also guide you at the time of our explanation.

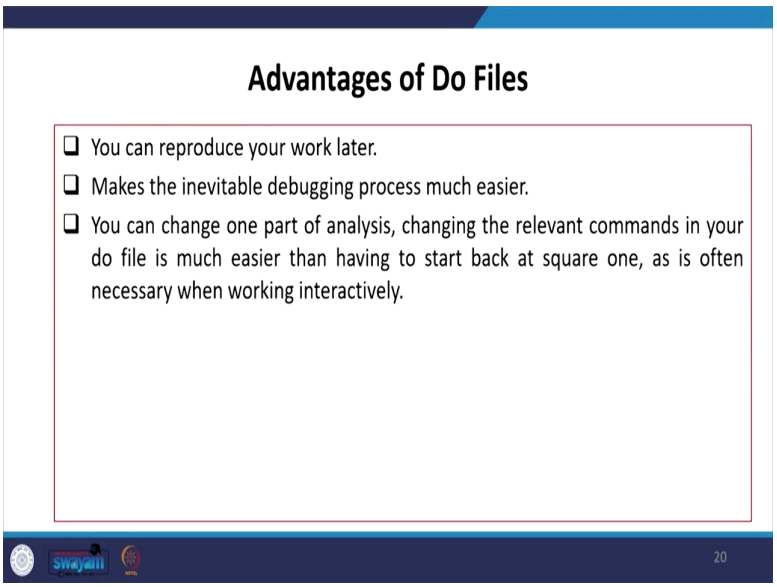
(Refer Slide Time: 43:04)



This is where the; once again we have shown it; how we can use our asterisk and forward slash marks; that will he help you to get the right command. And like here on the bottom; that is more important, I am just guiding it to like; suppose you have already given you a command in this line, then it ends at of triple forward slash; that on the next one we were actually using the weight.

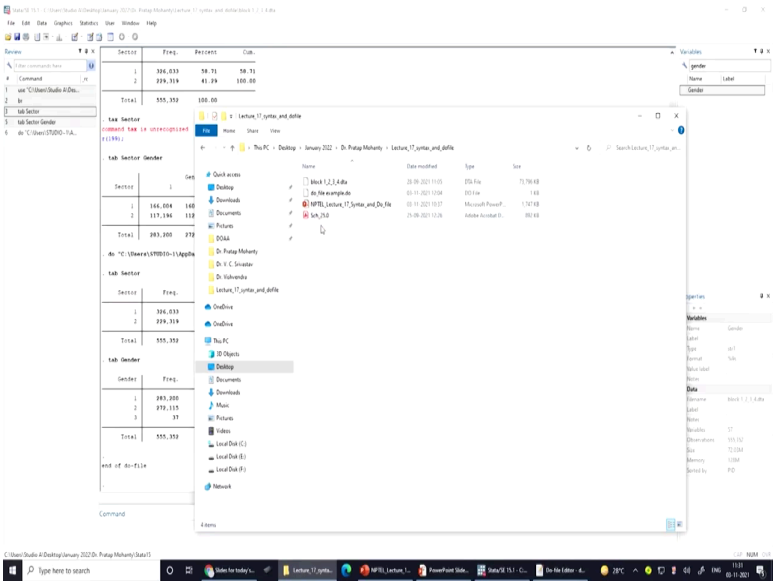
That means, this command; this command is actually utilizing this next line. Again, if you are using forward triple forward slash; this is again in fact, using other commands, as well alright.

(Refer Slide Time: 43:56)



So, what are the advantages of do-files? Do files there are lots of advantages; like you can reproduce your work later, which makes the in the inevitable debugging process much easier. You can change one part of the analysis, changing the relevant commands in your do-file is much easier and then having to start back at square one, an as is often necessary when working inter actively. So, we have already shown it some supplementary files; I can also show it once again, I am not going to spend much time; just going to show it once again.

(Refer Slide Time: 44:36)



On the screen; I have already explained a number of things; so do the file we are opening it, and some supplementary file; that will be also provided to you, at the time while you are

accessing all your files. I am just opening this folder once again which is going to be uploaded on your screen; just that folder. So, on the; 17 lecture we have these; so we all those things, we will actually give it to you.

So, we will keep it the sample data file, then do file example one, then the schedule one which I have repeatedly mentioned; 25 mentioned this indicating your health; schedule, 75 round; even other rounds of health care, those schedules 25 is also written. So, all those we are going to provide it for your help.

So, you need not worry at this moment, you just follow the material of NPTEL correctly, and you have all the options to download. I think I should not take much time, since we have already exceeded our limit for this lecture. So, I hope you will be enjoying; if you do your own exercise, these are all for today.

Thank you.