Handling Large-Scale Unit Level Data Using STATA Professor Pratap C. Mohanty Department of Humanities and Social Sciences Indian Institute of Technology, Roorkee Lecture 38 Pooled Cross-Sectional Data

Welcome friends once again to the NPTEL MOOC module on handling large scale unit level data using STATA. We are at the verge of understanding panel data. And, you have already been given the background of the panel data in the previous 2 lectures. In the last three lectures of the entire module, will be giving the hands-on experience of dealing with the panel data. So, this particular lecture is dealt with a form of panel, but clearly called as pooled cross-sectional data. So, with the meaning of pooled cross-sectional and panel and there are differences we already given. But some other details we are going to explain to you in this module.

(Refer Slide Time: 1:31)



So, when we think of panel or longitudinal data, it comes to our mind that we follow the same unit which gets repeated in different time period. This means that we will follow the same set of units of the observation over time, maybe of t1 period, t2 period and so on. While we discuss the full cross section, we take random samples in different time periods. So, our samples are completely random. It is not repeated. For example, when we take enterprise performance data or we take the performance data on enterprises. That is A, B, C in t1 and we will take the same

performance data on enterprises that is D, E and F in t2, then the enterprises we are referring not same. They are different in a different time period.

So, it is not repeated. The data are same, that means in terms of the information they are same, in terms of the variable they are same but the samples are different. Like we have different enterprises in our example and the key difference between pure panel data and pool cross-section is that the unit on which the data are collected is different. Units are to be compared. The fact that the random samples are collected independently of each other suggests that there need not be equal size.

The data can be pretty much analysed like ordinary cross-sectional data, except that we must use dummies, in order to differentiate the time period or the kind of changes in the enterprises over time. So, dummy method is going to be very useful in this setup. An important difference between pooled cross section and panel data is that with panel data, the effects represent changes over time within entities.

(Refer Slide Time: 03:48)

- □ An important difference between pooled cross sections and panel data is that with panel data the effects represent changes over time within entities.
- But in pooled cross section we can only talk about changes in population prevalence of attributes but we can not talk about individual changes.
- These kind of data mainly found in **survey datasets**.
- □ For many surveys, a cross-sectional sample is drawn **periodically**.
- □ These samples are drawn from the **same population**.

d one we can anky talk changes in nonvelation provelance of attribute

But in pooled one, we can only talk about changes in population prevalence of attributes, but we cannot talk about individual changes over time. These kind of data mainly found in survey datasets. And for many surveys, a cross-sectional sample is drawn periodically. These samples are drawn from the same population.

(Refer Slide Time: 04:09)

swayam 🛞

- Pooling of two cross section data make sense when the relationships among variables are unlikely to have changed notably.
- □ For pooled cross section data analysis we need to construct a time indicator variable (dummy variable) with value 0 and 1 or 1 and 2 and so on for period t₁, t₂...
- □ When it is of interest, a year indicator can also be interacted with another "x" variable of interest to examine whether its effect changed in that year compared to the other period(s) in the sample.

Pooling of two cross-section data makes sense when the relationship among variables are unlikely to have changed notably. So, if there are no such major changes expected, then pooling is generally preferred for pooling cross-sectional data analysis. We need to construct a time indicator variable that is dummy variable, let it be value 0 and 1 or 1 or 2 and so on by any binary form to define the dummy, to indicate time period 1 and time period 2 can be done.

When it is of interest a year indicator can also be interacted with another explanatory variable that is X of interest to examine whether its effect change in that year compared to other features. For example, in the Enterprises cases when we take another enterprise is taking loan in one period and taking loan in another period.

What is the difference in terms of enterprise performances in that case, the kind of loan or credit with the time factor interacted with the time factor has important interpretations? So, interaction dummy could be considered while making a pooled cross-sectional data over time.

(Refer Slide Time: 05:47)



Let us understand the construction of the pooled cross-sectional data. We are considering NSS data that provides cross-section data periodically. As we already discussed in NSS 67th. That was in 2010-11 and 73rd round of NSS of 2015-16 are going to be discussed now. In these two rounds, we are emphasizing on the unincorporated non-agricultural enterprises, which we already discussed in earlier modules as well. Especially we have the interest variables like enterprising types such as manufacturing, trade and other services. But make sure that these data do not consider the construction sector. So, coming to the survey, this is conducted in every 5 years.

(Refer Slide Time: 06:51)



And since the data provided by NSS is based on sample survey where in the same sample may not have been surveyed the next year. And the sample size may also vary from one year to another year. But the samples are drawn from the same population. What do you mean by same population? We wanted to mention that the population base, the enterprising base of concern for the survey were taken from the same districts in India.

So, the base of the population is same. But the random selection of the samples is different. The information collected in both rounds are almost same, especially relating to the variables of conscience. And besides that the 73rd round added some extra information, some of them we discussed earlier in our earlier modules.

(Refer Slide Time: 07:56)



So, coming to the nature of variable, as we already discussed that they are similar in both the datasets. So, pooling of datasets specially, requires the same variables almost in same order. So, new variable created is time and we wanted to capture the time effect of other variables after pooling the data.

(Refer Slide Time: 08:18)



Empirically, we wanted to explain this pooling of dataset using both the rounds of data. The objective of pooling would also be explained while explaining the pooling, the filtration and cleaning of data should also be done on the basis of the objectives we are targeting.

Here our objective is to understand the factors which affect the possibility of women being a necessity entrepreneur. So, we already clarified that whether necessity or opportunity-based entrepreneur, we are going to also suggest you one article which we published in the Standard ABDC ranked Journal, you can refer at the end. We are going to suggest that.

So, our interest here is to identify the important factors, alright the same example was also given during our lecture on Binary Choice model, especially on the week which we discuss about qualitative variables. Limited dependent variable models also. So, you please follow that and find out other clarification if you have desired to understand.

(Refer Slide Time: 09:46)



As the datasets are on enterprises, we first filter our data. For only women entrepreneur since our objective is to discuss on necessity women entrepreneur. So, we are emphasizing and filtering on that angle. The next important step is to get the variables of our interest which affect the women entrepreneur to be necessarily driven.

There is a note here, please mark it carefully, please make sure that the variables are with the same name in both the datasets. That is important variables should be with the same name,

number or observation may vary from one year to another because the datasets are survey data. And keep in mind that the samples are not the same in both the data sets.



(Refer Slide Time: 10:40)

Since different periods are considered. So, samples need not be same but variables with the same name must carry. As we already mentioned, that our dependent variable is women entrepreneur. So, enterprise type is our name of the variable so one is the code if it is of necessity type and 0 if it is opportunity type. The other variable that we already given emphasis of this aspect earlier like some continuous variables.

We have dummy variables in the model, we have categorical variables, we have some random variables that we mentioned, like we discussed about Enterprise age. And we also take the age square of that variable then from the dummy like sector dummy, mixed activity dummy and so on. Some categorical variable like social categories, bank accounts, etcetera. You go through and you will certainly find out how these variables affect the enterprise type.

(Refer Slide Time: 11:49)



We from the pool we can cross check and verify. And coming to the necessary steps involved in the construction of pooled data and their explanation, in this slide, we are mentioning the most important five steps. Launching the data in STATA. We start with 67th round data and then you can do on your own. But I am just telling you these important step and whichever is required direct operation and explanation we will use the software.

You simply open the 67th round data which already been guided earlier. It contains 40772 observations, there are women entrepreneurs and now filter the variables of your interest. And rename them like the variable which we already listed you can also change on your own, but try to give the name same in both the rounds. And create a new variable.

This variable is not there in the data we need to generate a variable that is here since two time period that is 67th and 73rd we are pooling, We are appending. So, only one variable is missing from the data. That is your time period. So, time we can give, it is equal to 0 generate a variable here equal to 0.

You can create it for the first period that is 67th round. We can save the dataset with the name pool 67th data will also show it and will also provide the sample of it for your easy

understanding. I think that is there I am not going to open it but I will just show it to you, like here.

(Refer Slide Time: 13:50)

	Computer	Removable Disk (G)						 +y Search 	Removable Disk (G)	P
0 🛛 - 🖷 - 🕺 - 🔒 🚽	Organize • Share with •	Rum New folder							11 - 71	0
T 0 2	and and and								- u	*
immands here	🖌 🚖 Favorites	Name		Date modifi	ed	Type	Size			- i k
and I w	Desktop	6th economic census		8/27/2020.4	:01 PM	File folder				
	👍 Downloads	1. binary choice model		10/6/2020 9	27 AM	File folder				E
r are no items to show.	💫 Recent Places	1. Factor analysis		9/8/2020 1:4	12 PM	File folder				r
		A BHDS2		8/27/2020 3	28 PM	File folder				
	🗸 🤤 Libraries	NSS73RD		8/27/2020 3	28 PM	File folder				- 11
	Documents	k panel data analysis		11/11/2020	1:15 PM	File folder				
	🛛 💰 Music	Regression		9/22/2020 3	36 PM	File folder				
	> 😣 Pictures	Reports		7/13/2020 1	0.30 PM	File folder				1
	🛛 🧸 Videos	Stata15		6/18/2020 5	58 PM	File folder				
		State wise Data IN ASCIL FORMAT			0.30 PM	File folder				
	🖌 💺 Computer	Supplementary DOCUMENTS		7/13/2020 1	0.31 PM	File folder				
	🗉 💩 Local Disk (C.)	EC6A ST05 UTTARAKHAND		6/8/2016 5:1	IZ AM	DAT File	22.327 KB			
	> qu Local Disk (D:)	EC6 Dictionary.dct		8/17/2020 1	231 A.	DCT File	1.08			
	> 🗇 Local Disk (E)	my de de		8/4/2020 2-2	IS PM	DO File	2.68			
) 🥪 Removable Disk 👌	FC6 uttacakhand extracted dta		8/17/2020.1	234 A	DTAFile	21.573 KR			
		Example data set dia		8/6/2020 11	R AM	DIAFile	17.754 KB			
	🗟 Network	eractice dataset BiDS dta		8/10/2020 5	01 AM	DTAFile	48.152.KB			
		practice dataset NSC3 block2-its		8/6/2020.2.2	N. PM	DTAFile	19.009 KP			
		Sa NSCLOW		8/12/2020.22	12 AM	Microsoft Office 6	2.388.87			
		(i) and		8/17/2020.2	JO AM	Managedt Office F	1.151.65			
		and the second s		4/17/2020.2	-97.0M	-recover once E.	1,231 50			
		Wonare 23- Neview of commands Solution and Solution		4/2//2020 2	-31% 10.04	-icrosort Office P.,	1.010 KB			
		ecture 20- ractol analysis		9/8/2020 32	10 PM	microsort Office P.,	2,243 KB			
	11 itamr	I Techure 78 - Linear Repression in STATA		9/17/2020.2	53.PM	Microsoft Office P_	7.586.08			
	55 nems									- 1
L	~									
-									 Notes 	_
									Variables	
									Observation	
(Command								* Size	
(2 🖸 🖌 🕯	D 🗿 🛄								- 1 0 (10
(° D) 🔺 (0 🛄								- 1 0 (12 14
	0								- 1 0 (
S1 Data Graphics State			1						- 2 0 (
S1 Data Graphics State	🖳 🇿 🔛	• Removable Disk (G) • panel data analysi	ġ					• 49 Searc	• 🔋 📀 (р Р
SI Data Guphica State E 🗟 : 🚊 : 🖉 : E	Computer Computer Computer Computer Share with	 Removable Disk (G) + parel deta analysi Dum New Kolder 	8					• 49 Searc	• 💈 🔷 4	р (р
SI Data Graphics Stat IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		Perrovable Disk (G) + parel data analysi Durn New folder Nears	is Date modified	Troe	Sar			• 49 Searc	 I I I I I I I I<td>ю М Р 0</td>	ю М Р 0
St Data Graphics Star IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII		Removable Disk (G) + parei deta analysi Bom New Kulder Name	is Date modified	Туре	Size			• 49 Sear	Donel data analysis Sec	р (р
SI Data Gaphica Star III III III IIII III III IIII IIII IIII IIII IIII IIIII IIIII IIIII IIIII IIIII IIIII			13 Date modified 10/24/2020 6:24 A.	Type DTA File	5ize 2.3/	19 KG		• 49 Sear	・ : ・ ・	р Р
ST Data Graphics State State State T 2 annuadi here una are no items to show	Organize + Computer Organize + Computer Organize + Share with feronites Doutrop Doutrop Doutrop Doutrop Doutrop Doutrop Doutrop	Removable Data (G) + genet data avaignt More Toble None Persong agend dia comproj undegrand dia	5 Date modified 10/24/2020 624 A. 10/24/2020 641 A.	Type DTA File DTA File	5ize 2,34 2,03	18 KD 17 KD		• +• Searc	・ : ・ ・	р Р О
SE Data Graphics Sont I T in it is it is a minute there G and Jc are no items to show	Dente Creatier - State with Creatier - State with Creations Dominant Sciences	Enmoster Dat.(2) + part data analys Enmoster Dat.(2) + part data Enmoster Dat.(2) Enmoster Dat.(2	8 Date modified 10/24/2020 6:14 A. 10/24/2020 6:43 A. 10/22/2020 3:21 A.	Type DTA Fåle DTA Fåle Adobe Acrobat D.	5ize 2,34 2,03 77	18 KB 17 K8 22 K0		• 4 See	A S O A	2 N
33 Data Gupha Sar	Computer Computer Computer Computer Computer Computer Computer Computer Commodit Second Acase		6 Date modified 10/24/2020 624 A. 10/22/2020 621 A. 10/22/2020 321 A. 10/24/2020 750 A.	Type DTA File DTA File Adobe Acrobat D DTA File	5ize 2.34 2.03 77 188,01	18 KB 17 KB 12 KB 19 KB	_	• 49 Searc	n panel dote analysis	
Still Deta: Guphes: State Image: State Image: State	Computer Computer Computer Computer Computer Computer Computer Computer Computer Common	Ennousite Data (G) • genet data analys Ennousite Data (G) • genet data Ennousite Data (G) • genet data example question data Gata Page Pricego genet data Pricedo genet data	B Date modified 10/24/2020 624 A. 10/24/2020 524 A. 10/24/2020 12 A. 10/24/2020 73 A. 1/2/2020 12 25 PM	Type DTA File DTA File Adobe Acrobat D DTA File DTA File	5/2e 2.34 2.01 77 188.01 191,30	88 KB 17 KB 17 KB 19 KB 19 KB		• 4• Searc	+ I ● ↓	р Р О
33 Dia Gupha Sar I S - <u>in</u> - <u>S</u> - <u>i</u> y 3 mounts here and <u>x</u> rare no items to show	Computer		8 Date modified 10/24/2020 6-14 A. 10/22/2020 5-14 A. 10/22/2020 3-21 A. 10/22/2020 3-21 A. 12/22/2020 3-25 PM 12/16/2019 2-52 PM	Type DTA File DTA File Adobe Acrobat D., DTA File DTA File DTA File	Size 2.34 2.01 77 188.01 191.34 103.70	81-00 17-74 17-24 19-04 19-04 19-04 19-04		• 49 Search	+ 1 • 4	
31 Dita Capha San III S • <u>ii</u> − <u>i</u> 7 annuab tere and <i>x</i> <i>x</i> are no terms to down	Compare C	Introvelia Data (G) + gener data analysi Introvelia Data (G) + gener data analysi Introvelia Construction of the	B Date modified 10/24/2020 614 A. 10/24/2020 613 A. 10/24/2020 324 A. 10/24/2020 750 A. 12/2020 750 A. 12/2020 750 A.	Type DTA File DTA File OTA File DTA File DTA File DTA File DTA File	5-2e 2.34 2.01 77 188,01 191,34 101,74 112,55	8 40 17 48 22 40 94 40 940 94 40 94 94 94 94 94 94 94 94 94 94 94 94 94		• 49 Sec	2 0 0	2 h
33 Data Graphic Son Son Son Son Son Son Son Son Son Son Son Son	Comparison		8 Date modified 10/24/2020 641 A. 10/24/2020 641 A. 10/27/2020 730 A. 10/27/2020 730 A. 10/27/2020 730 A. 10/27/2020 730 A. 10/27/2020 730 A.	Type DTA File DTA File DTA File DTA File DTA File DTA File DTA File	5.2e 2.34 2.01 77 188.01 191.34 103.74 112.55 33	44 48 47 48 27 48 49 48 49 48 49 48 49 48 49 48 49 49 49 49 49 49 49 49 49 49 49 49 49		• 49 See	pord tot onlys	
SI Des Gupha Ser S • ≜ • 3 • 3 mouada tiere vand ser no fierra to abox.	Computer	Introduct Data (G) + gener data analysi Inconsolin Data (G) + gener data analysi Inconsolin Data (G) + gener data analysi consequences analysis and data inconsolin data (G) + gener data	Oute modified 10/24/2020 624 A. 10/24/2020 624 A. 10/24/2020 613 A. 10/24/2020 7150 A. 10/24/2020 A. 1	Type DTA File Adde Ander Ander DTA File DTA File DTA File DTA File DTA File DTA File	5-22e 2.34 2.01 77 188.01 191.34 101.74 112.55 81 1,25	8 (0 17 (0 2 (0 0 (0 0 (0 0 (0 0 (0 0 0 (0 0 (0		• 49 Sec	2 0 0	
31 Det Graphic Ster S → d + S + S annuals here Q and x zero items to dow	Compare C	Intermeter Disk (C) + peer data analysi Termereter Disk (C) + peer data analysi Termereter gasent data reservery averages and as reservery averages avera	9 Date modified 10/24/0306-EX-A. 10/24/0306-EX-A. 10/24/0300-EX-A. 10/24/0300-EX-A. 10/24/0300-EX-A. 10/24/0300-EX-A. 10/24/0300-EX-A. 10/24/0300-EX-A. 10/24/0300-EX-A.	Type DTA For DTA For	5.2e 2.34 2.01 77 188.01 193.34 103.70 112.55 33 1.22 1.04	8 00 22 00 22 00 22 00 20 00 200000000		• 4 See	your dat angles	
33 Data Gupha Sara I S • <u>d</u> • <u>d</u> • 7 3 mouth ther 1 and <i>x</i> <i>x</i> <i>x</i> <i>x</i> <i>x</i> <i>x</i> <i>x</i> <i>x</i>	Compare Compa	Internetief Data (G) + gener data analysi Internetief Data (G) + gener data analysi Interneties analysis, analysi	Date modified 10/2/0006 6N A. 10/2/0000 6N A. 10/2/000 2N A. 10/2/000 2N A. 10/2/000 2N A. 10/2/000 2N A. 10/2/000 2N A. 5/10/2/00 2D A.	Type DTA File OTA File	5.2e 2.34 2.01 77 188.01 193.34 103.70 112.55 33 1.22 1.04 4.04	41 43 17 41 17 42 17 43 17 43 17 44 17 44 17 44 17 44 17 44 10 44 11 44		• 49 Seat	e 2 0 (
5) Det Granc See S • <u>a</u> + <u>S</u> • <u>a</u> mounts here or see to here to dow	Question Computer Question (c) Computer	Annovelin Dia (2) * peri data analysi Terre Video Terre Ter	8 Date modified 10/24/000 EVA A. 10/24/000 EVA A.	Type OTA Ne OTA Ne	Size 2.34 2.01 77 188.01 191.34 103.70 112.55 3.31 1.2X 1.04 44	4 4 4 7 7 4 7 2 4 7 2 4 7 3 7 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7 1 7		• 4: Sec	port dat onlys	
33 Date Graphic Son ■ 1 2 - 1 1 3 minution form 10 x rare no forms 10 abox	Compare C	Introduct Data (D) • gareti data analyo Introduct Data (D) • gareti data analyo Intro Textring agareti das Catagore, velopionistas Catagore, velopionistas Catagore, velopionistas Diolog, velotas Diolog, velotas pionol, Velotas	8 Date modified 102/00006424. 102/00006424. 102/000015441. 102/00001234 102/00001234 102/00001234 102/00001234 102/000012314 102/000012314 102/000012314	Type DTA for DTA for D	5/20 2.34 2.03 77 188.01 191.34 103.70 112.55 3.31 1.28 1.04 44	8 (d) 17 (d) 2 (d) 9 (d)		• • • • Sec	2 0 0	ka N ₽ ₽
A) Data Graphica San San Arris Angel San Arris Angel	Constant	Annovelin Dia (2) • peri das analys Inno New Kuto Nore recently peri das construction Nore recently peri da construction recently peri da construction Ployage period Ployage	8 Date modified 10/2/0008-124 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a. 10/2/0008-123 a.	Type DTA for OTA for O	Size 2,34 2,01 77 188,00 191,34 103,74 103,74 103,74 112,55 3,31 1,22 1,04 44	43 43 17 43 17 43 19 43 19 43 19 43 19 43 19 44 19 444		• 4 ₉ Sec	port data outjue	ka N ₽ ₽
31 Date Graphics Sore 11 1 2 1 11 1 2 1 11 1 2 1 11 1 2 1 11 1 2 1 12 1 1 1 13 1 1 1 14 1 1 1 14 1 1 1	Compare C	Internetief Data (G) + grand data analysi Internetief Data (G) + grand data analysi Internetief analysis analysis of the second s	Dute modified Date modified 10/24/0006424.a. 10/24/0006424.a. 10/24/0001421.a. 10/24/0001214.a. 10/24/0001214.a. 10/24/0001214.a. 10/24/0001214.a. 10/24/0001214.a.	Type DTA file OTA File Adde Antobe OL DTA file DTA file	522 2,34 2,01 19,134 10,374 112,55 3,1 1,20 1,04 44	8 4 7 21 72 4 9 24 9 24 9 24 9 24 9 24 9 24 9 24 9		• 49 See	por de organización de la consecutiva de la cons	
A) Deta Grupha Sar Sar 4 - 5 - 3 manata tere and K ran no dem to dos	Constant	Innovation Dia (C) + perior data analysi Ion Nore Matter Terrer	ContensionAnd Society 2010	Type DTA For OTA For OTA For OTA For DTA For OTA FOR O	5ce 2.34 2.03 17 188.05 191.34 103.74 112.55 31 1.20 1.04 40	81-01 17-03 17-03 17-03 19-03		• • • Ser	Loord Att angles	
Data Graphia Son Data Graphia Data Graphia Data Table	Compare Operator - Surventor - Operator - Surventor - Porteritor - Surventor - Porteritor -	Immovable Data (c) + geneti data analgo Immovable Data (c) + geneti data analgo Immovable Data (c) + geneti data analgo Immovable Data (c) + geneti data Immovable Data (c) + g	Date modified 1024/0308-144. 1024/0308-144. 1024/0308-124. 1024/0309-124. 1024/039-124. 1024/039-124. 1024/039-124. 1024/039-124. 1024/039-124.	Type DTA file OCA file OCA file OTA file	Size 2.34 2.01 77 188.01 19.1.34 10.3.74 11.2.55 3.31 1.2X 1.04 40	81.04 17.05		• 49 See	2 0 0	
A) Deta Crupha Sar Si • <u>in</u> • <u>f</u> • <u>i</u> mranab tree werd <u>x</u> are no items to abox	Constant	Innovation Dia (C) + perior data analysi Ion Nore Made Nore recensing used that	9 Date modified 10/2/0308-034.0. 10/2/0308-0. 10/2/0308-0.0. 10/2/0308-0.0. 10	Type DTA File DTA FIL	Size 2.34 2.01 77 188.07 191.34 103.74 112.55 31 12.55 31 1,22 1,04 44	8 04 7 08 7 24 9 04 9 04 9 04 9 04 9 04 9 04 9 04 9 0		• 4 ₈ Seec	2 Construction	20 h
Date Graphics for The Control Control Control Control The Control Control The Control Control The Control Control The Control	Compare Operator - Searement Operator - Searement Portugator - Searement Portugator - Searement Portugator Portugat	Introvestor Dola (C) + generi data analgo Ionn Nore Made Nore Remonstrating aport dia restrange aport dia	Date modified 1024/01084-04. 1024/01084-04. 1024/01084-04. 1024/01084-04. 1024/0184-04. 1024/0182-04. 1024/018-04. 1024/018-04. 1024/018-04. 1024/018-04. 10	Type DTA far Otta far Otta far Otta far DTA far DTA far DTA far DTA far DTA far DTA far DTA far Addre Annotat D.	Size 2.34 2.01 777 188.01 191.34 103.70 112.55 31 1.2X 1.04 40	0 0 0 17 0 17 0 17 0 17 0 19 0 19 0 19 0 19 0 19 0 19 0 19 0 19		• 4 Sec	port dat anajos	
A) Deta Crupha Der S • <u>in</u> + <u>S</u> • <u>i</u> T + mnaab tere weld x tere no dem to dow	Consure C	Innovatio Dia (3) * peri data antyti Nore Nore Terramity generi data rannyi generi data rannyi generi data rannyi generi data rannyi generi data Phity wategeneri d	Core modified 10/2/0006-824. 10/2/0006-824. 10/2/0000-824. 10/2/000-824. 10/2/000-824. 10/2/000-824. 10/2/000-824. 10/2/000-824. 10/2/000-824. 10/2/000-824. 10/2/000-824.	Type DTA for OTA for OTA for DTA for OTA for	5ce 234 201 191,34 101,70 112,59 31 12,59 31 1,20 40 40	11 43 17 43 17 43 17 43 17 44 19 44		• 4 Sec	porrel fatt analyse	
Dete Graphics Son The Constraint Son	Compare Operator - Searce - Compare Operator - Searce - Compare - Donne - Don	Annovativ Dok (C) + geneti data analgi Isan Nare Molor Nare Cash Nare Git Page Git Page Git Page Hoto, genetita Hoto, againet da Hoto, againet da Doto, Herca Doto, Herca Doto, Herca Doto, Herca Doto, Herca Doto, Herca Doto, Herca Doto, Herca Doto, Herca Doto, Herca Status Doto, Herca Doto, Herca Doto, Herca Status Stat	Date modified 1024/00084-04. 1024/00084-04. 1024/00084-04. 1024/0082108-01. 1024/0082108-00	Type DTA file DTA file	Size 2.34 2.01 191,34 103,74 112,55 3,31 1,22 1,24 4,44	8 4 7 5 7 5 7 5 7 5 7 5 7 5 7 5 7 5 7 5 7 5		• 4 Sec	port dat angle	
All Deta Crupha Der Sin tal Sin s Tie maab terr und tal Sin s terr terr to drem to drem	Consure C	Innovatio Dia (3) * peri data antysi Nore Nore Terremoti particità Nore Terremoti parti d'a resensito parti d'a	Core modified 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/0008-024. 10/2/008-024. 10/	Type DTA for OTA for OTA for DTA for OTA for	Sce 2.34 2.01 77 188.01 191.34 103.75 33 112.55 331 12.25 1.04 40	11 43 17 48 17 48 17 48 19 49 19 40 19 40 10 40 10 10 10 40 10 10 10 10 10 10 10 10 10 10 10 10 10		• • • Sorr		
30 Data Graphia Sor ↓ 2 · 2 via simulati fuer via ran no fermi Su show	Compare Operator - Searce - Compare Operator - Searce - Compare - Donney - Do	Interneted Data (G) • grand data analysi Interneted Data (G) • grand data analysi Interneted data analysi	8 Det modified 10/2/2010 6.24.4, 10/2/2010 6.41.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.64.4, 10/2/2010 7.27.14	Type DTA file DTA file	Sae 2.34 2.01 77 188.01 191.34 102.74 112.55 31 1.27 1.04 40	8 4 7 5 7 5 7 5 7 5 7 5 7 5 7 5 7 5 7 5 7 5		• (4) Sec	pore dat angle	
All Deta Crupha Der Sin ist - Sin S Tie mnaab terr und terr to drem to drem to drem to drem	Compare Constant - Source of Constant - Constant - Constant - Dones of Constant - Constant Constant Constant - Constant - Constant - Constant - Consta	Innovatio Dia (3) * peri data antys Nore Nor	Core modified 10/2/0006-024. 10/2/0006-024. 10/2/0000-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/00-024 10/2/	Type OTA fre OTA fre Aside Anabet D.	5/2e 2.34 2.01 77 198.34 193.34 193.34 103.70 112.55 33 1.22 10,70 40 40	0 (0) 7 (0) 7 (2) 9 (0) 9 (0)		• • • Sorr		
31 Date Graphic Sor ↓ 2 - 2 v 3 minution free ↓ v 4 v 4 v 4 v 4 v 5 v 5 v 5 v 5 v 5 v 5 v 5 v 5	Compare C	Introduct Data (3) • panel data analysi Inter Note Materia	8 Determotified 10/24/2004 644 A. 10/24/2004 644 A. 10/24/2004 546 A.	Type DTA File DTA File DTA Afte DTA file DTA fil	Size 2.34 2.01 191,34 103,74 103,74 104,74 1	11 40 7 44 7 2 48 7 9 48 7 9 48 7 9 48 7 9 48 7 1 4 7 1 1 1 1		• 4 Jac	pore dat angle	
A) Deta Crupho Der S • da - S • 2 T + mnacher C web × tere no dem to dow	Constant	Innovatio Dia (3) * peri data antysi Nore Nore Terremity guerd fa reampt guerd fa Nore Nor	Core modified 19/2/0306-034. 19/2/0306-034. 19/2/0300-034. 19/2/030-034.	Type OTA fre OTA fre Adde Anyber D.	522e 2.34 200 191.34 103.74 112.55 331 1.22 1.04 40	0 0 7 0 7 0 9 0 9 0 9 0 9 0 1 9 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		• 4 Sec		
31 Date Graphic Sor ↓ S + 4 + 5 + 3 T 3 minution from ↓ 4 rare no from Sta show	Compare C	Introvetier Date (G) + gener data analysi Inter Note (Gate Note (Gat	8 Det modifiel 10/2000 644. 10/2000 645. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758.	Type DTA File DTA File DTA Arte DTA Arte	Size 2,34 2,010 19,1,34 10,1,74 11,255 3,1 1,22 1,04 4,0	11 40 7 48 7 2 48 7 2 48 7 9 48 7 9 48 7 4 4 7 4 4 7 4 4 7 4 4 7 4 8 7 4 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 7 8 7 7 8 7 7 8 7 7 7 8 7		• 4 Sec		
A) Deta Crupha Der Sin tal Sin tal minada tere wei ∠ zeren berna te done	Constant	Innovatio Dia (3) * peri data antys Ion Nore Terrenits guerd fa reampt guerd fa r	Care modified 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-814.a. 19/2/0306	Type OTA fre OTA fre OTA fre OTA fre OTA fre OTA fre OTA fre OTA fre OTA fre Aside Arroba D.	5/2e 2.34 201 191,34 101,94 102,95 1,28 1,28 1,28 1,28 1,28 1,28 1,28 1,28	0 0 7 0 7 0 9		• 4 Sec		
31 Date Graphic Sor ■ Sor - 2 - 3 T 3 minution from 10 x rate in the story down	Compare C	Introvetier Date (G) + gener data analysi Inter Note (G) Norm Norm Comparison of the second seco	8 Det modifiel 10/2000 644. 10/2000 645. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758.	Type DTA File DTA File DTA Arte DTA Arte Assider Annibell D.	52e 234 2010 1913 103,70 1913 11255 33 11255 33 1255 34 1040 44	11 40 7 48 7 2 48 7 2 48 7 9 48 7 9 48 7 4 4 7 4 4 7 4 4 7 4 4 7 4 8 7 4 8 7 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7 8 7				
All Deta Crupha Der Sin - All - Sin Tit annanda terr annanda terr terr no dem to dow	Compare Constance	Innovatio Dia (3) * peri data antysi Ion Nore Terremiti guerd fa reampt guerd fa	Core modified 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a. 19/2/0306-824.a.	Type OTA file OTA fil	Size 2.34 201 77 1880 1910 1910 1910 1925 104 40	0 0 7 0 7 0 9 0 9 0 9 0 9 0 1 9 0 1 9 0 1 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0		• 4 Ees	Partier Company	
3) Date Graphics Sore 11 Total Graphics Sore 12 Total Total Total 13 Total Total Total 14 Total Total Total 15 Total Total Total 16 Total Total Total 17 Total Total Total 18 Total Total Total	Compare C	Introvetie Data (3) + gener data analysi Inter Note Materia	8 Det modifiel 10/2000 644. 10/2000 645. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758. 10/2000 758.	Type DTA File DTA File DTA Arte DTA Arte	502 2.31 201 77 188.07 191.34 103.24 104 40 40	11 40 7 44 7 2 40 7 9 40 7 9 40 7 9 40 7 4 40 7 40 8 40 7 4 40 7 40 8 40 7 4 40 7 40 8 40 7 4 40 7 40 8 40 7 4 40 7 40 7			Paraden Control Contro Control Control Control Control Control Co	
All Deta Coupina Der III - La - S - C T - I minada terr varia - zer no dem to don	Compare Constant - Searceit Constant - Searceit Constant - Searceit Constant - Searceit Constant - Searceit Constant Constant - Searceit Constant Consta	Innovatio Dia (3) * peri data antysi Ion Nore Terremiti guerd fa reampt guerd fa	Core modified 19/2/0306-824.a. 19/2/0306	Type OTA for OTA for	Size 2.34 2.01 77 198.34 199.34 103.74 11255 33 34 120 120 44	0 0 7 0 7 0 9 0 9 0 9 0 9 0 1 9 0 1 9 0 1 9 0 9 0 1 9 0 1 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0		• • •	Partier P	

So, it is there now pool 67 will provide this, and we also saw it once. Do you pool 73rd. So, that is going to be explained to you shortly? I am just giving the clarification of the concepts.

(Refer Slide Time: 14:10)



Similarly, we open the another round data that is seven third and we again repeat all those five steps that have been already mentioned here now and then two to five steps. That is very important. And to mention here that after filtration the women enterprises here for us for the explanation is of 34930. And we again generate the same year as the variable name to be one, because this is another time period data. And will save with the pool 73rd data.

We are not following the above steps as we have already filtered out our data, the data which we have kept here for recognition, we already done all those steps. So, we are not repeating on each step that we have mentioned. Onwards we are going to experiment in front you like we are opening the data that is pooled data of 67th round. Let me open that first for you.

(Refer Slide Time: 15:24)



It is here now I am just going to open it, it is here. So, the pooled 67th data, as I told you it has already been opened now. what I will do.

(Refer Slide Time: 15:43)



I will go by, just read the red colour for the text. That is for preparing a pooled dataset based on two different rounds of NSS. We need to append those datasets like two data sets. We are discussing 67th 73rd we need to append, having the exactly same variables then only it will simply append and basically it is a vertical appending procedure. The observation of 67th will be from the beginning and at the end part the 73rd observation will be simply appended or added vertically. Append is the right word add sometimes you may infer that it is a numerical operation, but we are simply trying to say that it gets kept in different rules of the data.

(Refer Slide Time: 16:33)



Also everything right now. So, what we are doing, we are trying to show you the append.

(Refer Slide Time: 16:41)

та	x	* Variables
		▲ Filter variables here
mand been		Name Label
"Gloanel data analysist.	Statistics/Data Analysis StataCorp	sectornum Sector
	4905 Lakeway Drive	mix activity Was pursuing
	Special Edition College Station, Texas 77045 USA	social catego. Social group
	970-554.4400 stata.com	enterprise ty. Enterprise Typ
	979-696-4601 (fax)	nature opera Nature of op
		prbim facd Whether prot
	20-studier flata lab propertual license	assistance rc Whether assis
	Incial summer distribution	growth status. Status of enter
	Licensed to: IIT Roorkee	activity group
		trad sec
	Notes:	enterprise age
	 Unicode is supported; see help unicode_advice. 	southstates RECODE of st
	Maximum number of variables is set to 5000; see help set maxwar.	location entr. RECODE of ic
	Constitution of the control of the data	
	. use .o.: (panel data analysis (pool_e/th.dta.	Properties
		B
		Variables
		Label
		Type
		Format Michae Labert
		Note:
		f Data
		Filename pool 67th
		 Notes
	halo cat manor	Variables 15
	nay styleses	Observations 40,772
	Command	Size 1.17M
		Memory 64M
		Sorted by

Since we have already used the data 67th to open. So, you are opening append using 73rd data. The location of it is here.

(Refer Slide Time: 17:02)



. (0 D	33	۲.	Da -													
	sectorina	m[1]															
	Aectornum	154.40	tvity	social_cat-y	encerpris-pe	nature_ope-n	prbin_fecd	Assistance-d	growth_sta-s	activity_g-p	tred_sec	enterpris-pe	southstates	location_e-e	age_equare *	Variables	
1	rural	1	360	Others	1	perenalal	54	80	1	nanfet	traditional	21	1	LASI della	441 =	A Elter analle	
2	vrben		36.5	50	1	perennial	Tes	80		trade	nontreditional	2	2	outsideRt	•		
3	rurel.		18-5	080	1	perennial	50	80		others	treditional	1	1	inside/II	1	Marie	Label
4	urben		35a	080	1	perennial	Tes	84	1	matfct	treditional	21	1	inside/II	441	🖌 sectornum	Sector
5	vrben		80	Others	1	perennial	80	Ro.		matfct	treditione)	3	1	stasdeltt		M majactivity	Was pursuing m
6	rurel		345	080	1	perennial	Tee	84		nentes	treditional	2	1	inside/ft	•	Social catego.	. Social group of
,	urben		80	080	1	perennial	No	35	2	nanfct	treditional	21	1	sn#sdelf#	441	enterprise_ty.	Enterprise Type
	roral		3L9	080	1	perennial	34	30		nant or	creditional	1	1	10.41deRI	1	M nature_opera.	Nature of opera
9	urben		No	080	1	perennial	50	50	2	trade	nontreditional	3	1	insideRt		Prblm_facd	Whether proble
10	rarel		340		1	perennial	Tes	No		trade	nontreditionel	3	1	outsidelt!	,	B assistance_rc.	Whether assista
11	rural		36 0	57	1	perennial	No	80	2	BALFCS	treditional	51	2	stasdelft.	2601	growth_status	Status of enterp
12	rurel		150	080	1	perennial	Tes	80	2	manfes	treditional	,	1	snasdelfit	49		
13	rurel		80	Others	1	perennial	80	86		BALFOL	treditional	2	2	15#1deRt	•	🕑 trad_sec	
14	ratel		80	080	1	perennial	No	80		nantes	treditional	1	1	steldeltt	1	enterprise_age	
15	rural		340	8	1	perennial	Tes	Tes	3	BATEC	treditional	41	1	1581deRR	1681	Southstates	RECODE of stat
16	urben		80	080	1	perennial	Tes	34		nanf ct.	treditional	2	2	inside/R		location_entr.	RECODE of loca
17	urben		×0	50	1	perennial	No	No		trade	montreditions]		1	outsidelli	16	vacuores snapt	103
18	rurel		340	Others	1	seasonal	50	80	1	manfet	treditional	•	2	staideRE	14	Properties	
19	rerel		360	57	1	perennial	Tes	80	2	nation	treditional	29	2	outsidelfi	841	4 Variables	
20	urben		80	080	1	percential	Tes	80	3	netter	treditional	,	2	15.81-delitt	81	Name	sectorium
23	vrben		80	080	1	perennial	Tes	80		netes	treditional	2	2	insidelli	•	Label	Sector
22	rurel		80	Others	1	perennial	Tes	30	2	manfet	montreditionel	21	1	11#1del01	441	Type	byte
23	urben		80	080	1	perennial	Tes	85	2	BALECS	treditional	14	1	ineidelti	196	Format	%10.0g
24	rurel		No.	57	1	perennial	Tee	Tes	2	netfer	treditional	31	1	11#1delfi	941	Value label	sectornum
25	rurel		350	080	1	perennial	So	80	2	nanfct	nontreditionel	31	1	outside/II	961	Notes	
26	rurel		¥0	080	0	perennial	Tee	80	1	nation	treditional	13	1	outsidelli	121	* Data	
27	uthen		80	50	1	perennial	80	80	1	manfin	treditionel	6	1	snesdelf.	36	Elename	pool_67th.dt
2.0	urben		360	080	1	perennial	No	80	2	nenfox	treditionel	11	2	inside/0	121	Label	
29	urben		No.	80	1	perennial	No	84		others	treditional	1	2	outesdellit	1		
30	ratel		30	8	1	perennial	Tee	Tes	3	nanf ct.	treditional	51	1	insidelli	2601	Vanables	
31	urben		350	Others	1	perennial	80	80		manfes	traditional	2	2	snesdelf#			
			11.0	dahaaa	11		V.a.	-						14.41.4.175		2/26	2.375
-															,		

We are copying the location of it. 67 already opened, we are simply we are adding the 73^{rd} round. 73^{rd} so it is here. Alright. We have appended the data. You can check the number of the observation recently the case has been added, increased to 75702. And the variables are same that those are 15 we have carried from the 67^{th} round and since the number of variables are of 15 in both the rounds. So, number of variables remain the same.

And our observation increased by another the number of observations in the next round. So, these are the data we have added. And we have appended now just to show you that your variable have already created and we have deliberately made it a dummy one to identify that zero indicates your 67th round data and whenever one is there, one indicates you a 73rd round data.

So, suppose I just wanted to mix like age of the enterprise times year effect of it, or location of the enterprise. And year effect we can simply multiply this data. I can find out the location with effect on the dependent variable so there is a better interpretation possible in our ppt.

(Refer Slide Time: 19:04)



Let me go through the exact steps that we have already followed. After pooling one thing for sure, we have to save it, save the data with a convenient name that can identify pooled 67^{th} as well as 73^{rd} dot dta will come by default since dta file has already been opened. So, you please take a name with a convenient identification 67 and 73 both are shown here. You can take any other name, but for your clarification, we are just trying to keep some suggestions.

So, as we mentioned the number of observation has been increased already after pooling, alright. So, analysis can be done, we can do n number of explanation but some just sample we have done those explanations many times likewise we do in the earlier statistical operations. (Refer Slide Time: 20:05)



We can simply do the same thing we can do describe. Simply describe, can clarify. It gives the information about the kind of storage the byte or float or numeric or string. And also the value label as well as variable label can also be clarified.

(Refer Slide Time: 20:32)



Similarly, the 15 variables, we have already appended. So, the 15 variable information will come.

(Refer Slide Time: 20:42)



Similarly, you can do summarize, you can just go through on your own. But I am not operating here much. But I think it is too simple to understand, just to mention that since our dependent variable is binary in nature, because we already pointed out that it is whether the enterprise is of

necessity type or opportunity type. So, we can apply either of the 3 models we discuss in the binary choice model LPM model or logit or profit model.

So, you go to that section once again and verify and I am sure you will enjoy going through the data. In our binary response model, we have already check the appropriate model and it will be very useful for you. The same model we can apply to do it here.

(Refer Slide Time: 21:37)



We can operate it here just for your another round of confidence. Let me do it. Let me just do it once to that particular model. And just for your confidence, only though usually we suggested logit or profit based on some logic and we will once again operate here.

(Refer Slide Time: 22:00)

w Tax	obs: 7	5,702				Variables	•
liter commands here	vars: size: 2.21	15			19 Sep 2018 20:13	A Filter variable	is here
Command _rc -		.,				Name	Label
use "G\panel data analysis\	3	storage	display	value		sectornum	Sector
append using "G1panel dat	variable name	type	format	label	variable label	mix activity	Was pursuing mic
des						social catego	Social group of a
	ectornum	byte	810.0g	sectornum	testes	entermise by	Enternoise Type /
	is estuitu	buta	\$10.0x	niv actuit	sector	entre coera	Nature of operat
	ax_accord	byce	810.0g	111 BCCC1	Van murauing mixed activitu?	intere operation	Whether proble
	ocial category	, byte		social ca	Peory in the second of the sec	proim_tacd	
-		2700			Social group of owner	assistance_rc	Whether assistan
	interprise type	byte	\$10.0g		Enterprise Type during last 365 days	growth_status	Status of enterp
	ature operat-r	byte	\$10.0g	nature op	eration	activity_group	
		,	,		Nature of operation	trad_sec	
-	orbin_facd	byte	\$10.0g	prblm fac	4	enterprise age	
-	-				Whether problems faced last 365 day	southstates	RECODE of state
	ssistance_rovd	byte	%10.0g	assistanc	e rcvd	location entr	85000 of lock
					Whether assistance recd last 365 day	incaron (entr.	1000101004
4	rowth_status	byte	%10.0g		Status of enterprise	Properties	
	ctivity_group	float	89.0g	.0g activity_	group	8	
	trad_sec					+ Variables	
1		float	\$14.0g	trad_sec			
•	interprise_age	float	89.0g 89.0g			E Label	
	outhstates	Dyte			RECODE of statenum (State code)	Liber	
,	location_entr~e	byte	89.0g	location	entrprise	Type	
		0	40.0-		RECOUR of location (Location of enterprise)	Format	
	de sdrate	float	49.00				
1		11040	17.19			Notes	
1	forted by:					4 Data	
	Note: Data	ant has	changed at	ince last as	red.	Filename	pool_67th.dta
						Label	
						Notes	
						Vanables	
						Observations	
c	ommand					Size	2.17M
1	ogit enterprise	11					

With the same data and the result is in front of your screen. It is I think walking now the result has come. Alright. So accordingly, different variables and the interpretation everything is there and our dependent variable is enterprise type. As I told you age and by square of that age whether it is significant or not. And about pseudo r square is not that important. We already discussed during our lecture you go through and surely you can able to interpret it. So, let me proceed further.

(Refer Slide Time: 22:43)



Coming to that time dummy we have already discussed that. We can have interaction dummy also. We wanted to get the change in relationship between outcome and predicted variable through time. Basically, wherever you multiply the time effect, it will give important information. But since dummy has already been introduced, so the degrees of freedom gets reduced in the model. So, we need to be careful about that while analysing the dummies.

(Refer Slide Time: 23:24)



Most importantly, if you wanted to understand the model very correctly and the interpretation, everything is there in this link. This is our article published in Global Business Review in the year 2020. So, you download and read. I am sure you can find better interpretation and a systematic interpretation of it with this. I thank you all, will now look forward to your involvement for the next lecture because next lecture we are going to give you a systematic guidance on fixed effect, panel effect discussions. Thank you so much.